Detection of antiretroviral drug resistant minority variants in patients failing second line antiretroviral therapy in South Africa

Mhlekazi Cathrine Molatoli

A dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of

Master of Science, January 2019

Declaration

I, Mhlekazi Cathrine Molatoli, 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.

_____

_____day of_____May_____2019

Acknowledgements

I dedicate this dissertation to my late Mother, who was always my rock throughout the years. Her emotionally and financial supportive has gotten past the last few years. I will forever be grateful for all she has taught me and will always love her.

To my supervisor, Prof Maria Papathanasopoulos, thank you for granting me the opportunity to be a part of your research team and trusting in my capabilities. I have learned and improved not only on my thinking skills but also my writing skills under your guidance. I wish to continue to learn from you over the coming years.

To my co-supervisor Dr Kim Steegen, you have gone over and above for me. Thank you from the bottom of my heart for the patience, guidance and support you have shown and continue to show me not only as my supervisor but a mentor.

I would also like to give a huge thank you to the NICD team of Dr Gillian Hunt, Johanna Ledwaba and Dr Mukhulid Yousif. Thank you for your hospitality, taking the time out to our busy schedules to always accommodate me and my million questions and editing and commenting on my abstracts. A special thank you to Johanna, for having the patience to help me with Geneious, you are the best teacher.

To the NICD NGS core team, thank you for your guidance and training through Next generation sequencing, which was a completely new field and technology.I hope to continue to learn from you all as I hopefully find my feet in this field.

To the NHLS genotyping team, Dr Irene Ketseoglou, Vibha Kana, Esrom and Lindiwe, thank you for welcoming me into your team over the last two years. Thank you for your advice, training and introduction to the HIV genotyping world.

To my friends and colleges at the HIV Pathogenesis Research Unit, thank you for creating a positive environment that stimulates interesting thoughts and discussions.

Lastly, I would like to also give a special thank you to Mrs Nancy Tumba, for all the advice, guidance, mentorship and willingness to help.

Abstract

The development of antiretroviral (ARV) drug resistance to Protease (PR) Inhibitors (PIs) requires the emergence of both primary and secondary/ compensatory mutations in the PR enzyme. Several studies have shown that most patients virologically failing a PI based ARV drug regimen do not harbour PI drug resistant mutations (DRMs). The presence of minority HIV-1 PI variants undetected by population-based Sanger sequencing (sensitivity limitation of ≥ 15- 20%), have been hypothesised to possibily contribute to the development of PI resistance in these patients**.**

A total of 188 participant samples previously collected from six provinces across South Africa and virologically failing a PI based second-line regimen were available for this study. Population based Sanger sequences and associated DRMs were available for all the participants. Following viral RNA extraction using the EasyMag Extraction Kit (Biomerieux, INC, France) and the NucliSeNS® EasyMag instrument (Biomerieux INC, France), RT-PCR amplification and sequencing-by-synthesis on the next generation sequencing (NGS) platform, Illumina Miseq (California, USA), was successful for 158 of the participants. Identification of all HIV-1 DRM variants was performed using both Deepchek® and Geneious® analysis tools.

Following a comparison of DR variants obtained using Deepchek® and Geneious®, a minimal cut-off of ≥ 4.5% was used for the remainder of the study. NGS detected virus containing PI DRMs in 26 of the participants versus 24 participants detected by Sanger sequencing. Minority HIV-1 variants as defined at ≥ 4.5- 20.0%, were identified in viral PR for five of the 26 participants sequences by NGS and resulted in PI regimen drug resistance for only two of the five participants. Overall, NGS detected PI DRMs resulting in PI resistance for 20 participants versus 18 participants by Sanger sequencing. Thus, using NGS based genotyping results, an additional two participants would be referred to switch to a third-line ART regimen. Furthermore, the detection of minority RT inhibitor variants in three of the abovementioned 20 participant sequences resulted in a further change in the recommended third-line RT backbone.

Our findings confirm previous studies and show that the majority (83.5%) of study participants that were failing a PI based regimen do so in the absence of detectable, known PI DRMs. It is likely that non-adherence to the PI based regimen may explain the observed virological failures in the absence of relevant DRMs. Alternatively; an as of yet unidentified mechanism of DR may contribute to changes in PI drug resistance or susceptibility.

Future work must focus on elucidating whether additional viral mechanisms or patient non-adherence are responsible for virologic failure to a PI based regimen in the absence of PI DRMs.

Table of Contents

List of figures

List of tables

List of abbreviations

3TC	Lamivudine
ABC	Abacavir
AIDS	Acquired Immune Deficiency 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
CDC	Centres for Disease Control and Prevention
CD4	Cluster of Differentiation 4
d4T	Stavudine
ddi	Didanosine
DRM	Drug Resistance Mutation
DRV	Darunavir
DTG	Dolutegravir
EFV	Efavirenz
EVG	Elvitegravir
ETR	Etravirine
FDA	US Food and Drug Administration
FTC	Emtricitabine
GB	Gigabyte
HAART	Highly active antiretroviral therapy
HIV	Human Immunodeficiency Virus
IN	Integrase
INSTI	Integrase strand transfer inhibitor
PCP	*Pneumocystis Carinii* Pneumonia
pH	*Potential of Hydrogen*

PIC	*Pre integration complex*
pol	*Polymerase*
pro	*Protease*
RAL	Raltegravir
RNA	Ribonucleic acid
RT	Reverse transcriptase
RT-PCR	*Reverse Transcription Polymerase Chain Reaction*
RTV	Ritonavir
SBS	sequencing by synthesis
SIV	Simian immunodeficiency virus
SPRI	Solid Phase Reverse Immobilisation
SQV	Saquinavir
LPV	Lopinavir
NFV	Nelfinavir
NGS	Next Generation Sequencing
nNRTIs	Non-nucleotide Reverse Transcriptase Inhibitors
NRTIs	Nucleoside Reverse Transcriptase Inhibitors
NtRTIs	Nucleotide Reverse Transcriptase Inhibitors
NVP	Nevirapine
NT	Neutralize Tagment
TD	Tagment DNA Buffer
TDF	Tenofovir
WHO	World Health Organisation

Chapter 1

1. Introduction

Since the first officially reported cases in 1981 (1), infection with human immunodeficiency virus (HIV), the etiological agent of Acquired Immunodeficiency Syndrome (AIDS) (2), remains the most devastating human virus in our history. The World Health Organisation (WHO) in 2017, estimated that 36.9 million people were globally living with the virus and of this, 940 000 died from AIDS related illnesses (3). As illustrated in Figure 1.1, there has been a steady decrease in AIDS related deaths with, for example, an estimated 1.7 million deaths reported in 2010 and 1 million in 2016. These substantial reductions in AIDS associated morbidity and mortality can be attributed to the development and introduction of Antiretroviral drugs (ARVs) disseminated in the form of highly active antiretroviral therapy (HAART) from 1996. There have been global increases in the number of HIV infected people initiating HAART, and moreover more potent and/or novel ARVs have been developed over time.

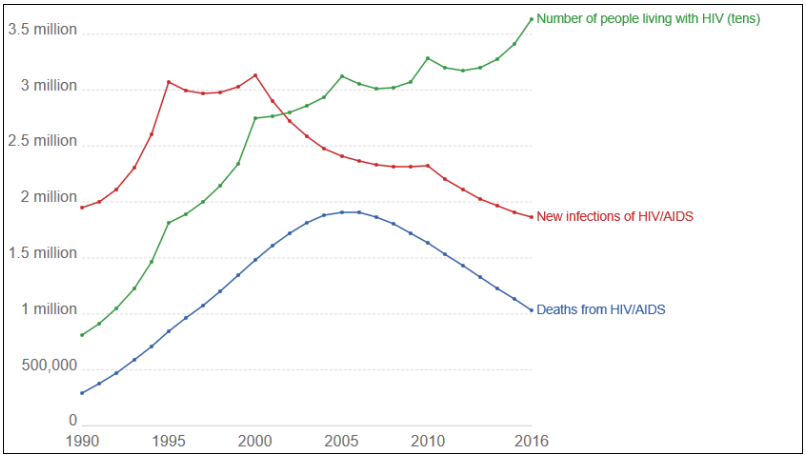

Figure 1.1: Total number of HIV/AIDS infected people, new infections per year and total number of AIDS related deaths since 1990 to 2016, globally. Copied from Our world in data (https://ourworldindata.org/hiv-aids).

1.1. Brief History of Human Immunodeficiency Virus

There is evidence supporting the emergence of HIV/AIDS in humans from as early 1908-1931 (4), although the cluster of diseases was only officially recognised in 1981 by the Centres for Disease Control and Prevention (CDC) in the United States of America (USA) (5). This recognition followed the reporting of increased incidences (6) of

previously rare diseases of *Pneumocystis Carinii* Pneumonia and Kaposi's Sarcoma (KPS), which were usually associated with both severely immunosuppressed patients (7) and with multi microbial infections (8). The first officially reported cases involving this group of disease entities involved five young previously healthy homosexual and bisexual male patients in Los Angeles, USA. Subsequently, the CDC continued to receive an influx of reports involving several numbers of homosexual men succumbing to this group of disease entities.

In the same year, a retrovirus was identified as the etiological agent of AIDS, through the works of American Robert Gallo (9) who named it Lymphadenopathy Associated Virus (LAV) highlighting its isolation from lymph node biopsy, and a French research group led by Luc Montagnier (2, 10), who identified Human T-Lymphotropic virus III (HTLV-III) whose genomic structure resembled that of HTLV-I to also be the causative agent of AIDS. The viral isolates from the two groups were identified to be the same virus following whole genome sequencing, and the virus was then officially renamed Human Immunodeficiency Virus. Over the coming years, more discoveries led to a better understanding of the virus's ancestry, structure and epidemiology.

1.2.HIV epidemiology

There are two genetically distinct types of HIV, HIV type 1 (HIV-1) and HIV type 2 (HIV-2) which constitute the pandemic form of the virus. HIV is a zoonotic disease. Evidence pointing to the exact time of this zoonotic transmission is not as concrete, but it is highly possible, that it was during the bush hunting and wild meat consumption of the late 19th to early 20th century (11), a practice which is still seen in parts of Africa today. The practice have been linked to other Zoonotic infections such as foamy viruses in chimpanzees (12). Simian immunodeficiency virus (SIV) from chimpanzee subspecies *Pan troglodytes* (SIVcpz) crossed the species barrier, resulting in HIV-1 (13), while SIV from sooty mangabeys subspecies *Cercocebusatys* (SIVSM) resulted in HIV-2 (14). The two virus types share a few similar clinical, biological and genetic similarities (15).

The two types of virus share a similar gene arrangement, with the exception of accessory gene, *Vpx* which is only present in HIV-2 and *Vpu* in HIV-1(16). Biologically the virus types share similar intercellular mechanisms of replication, target the same cell types and use both the same receptor and co-receptors for virus entry (17). They share

the same modes of transmission which includes fluid exchange through sexual activity, needle sharing and blood transfusions. By contrast, disease progression to AIDS is slower with HIV-2 infection, however the progression to an advanced stage of the disease presents with manifestations which include but are not limited to the development of opportunistic infections, tuberculosis, Central nervous system infection and others. Despite such similarities the two types of the virus remain distinctly different, in their pathogenesis resulting in significant differences in their diversity, distribution and physio-pathological characteristics (13, 18, 19).

There are fewer reported cases of HIV-2 globally, and the virus is largely confined in West Africa (20, 21) and imported into parts of India and Europe (22), with reports of its decline in prevalence (23, 24). On the other hand, HIV-1, the pandemic form of the virus shows large diversity and infection rates. HIV-1 comprises of 4 groups; M (major); N (new or non-M, non-O) (25); O (outlier) (26) and group P (27) as depicted in Figure 1.2. HIV-1, Group M (28) is the most predominant, and exhibits high genetic diversity based on phylogenetic analysis of multiple sub-genomic regions mainly; *env, pol* and *gag* gene sequences. There has been an estimated 25-30% amino acid variation reported among the subtypes themselves, with an additional estimated 15-20% variation within each subtype (29). This level of diversity has given rise to nine distinct subtype and sub-subtype lineages mainly A-D, F-H, J and K (27, 30) of which subtype C accounts for more than 50% of the worldwide infections. In addition, an individual can be dually or multi subtype infected, resulting in the recombination of genetic material to create circulating recombinant forms (CRFs) and unique recombination forms (URFs) (31).More than 90 CRF'S have currently been identified (32).

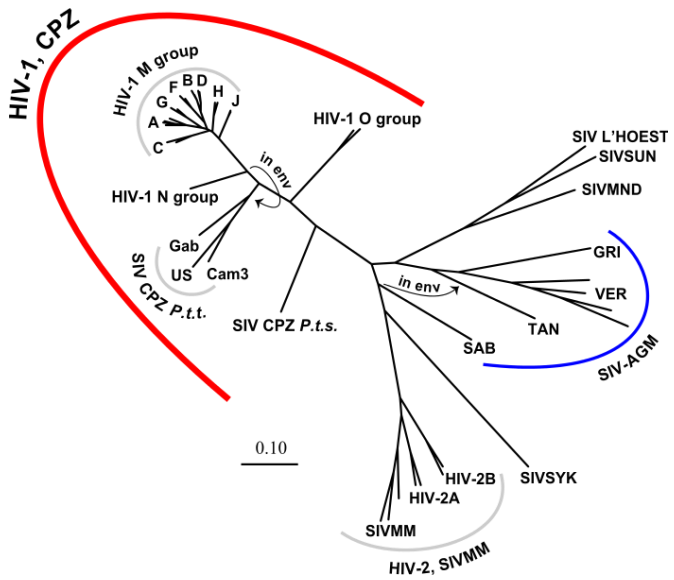

Figure 1. 2: Phylogenetic maximum likelihood tree constructed utilising *pol* gene sequence, showing HIV diversity and the ancestral relationship of HIV-1, HIV-2 and Simian Immunodeficiency Virus (SIV). Copied from Korber *et al.,* 2000 (33)**.**

1.3. HIV-1 genome and structure

The deduction of the HIV-1 structure, genome, viral life cycle and disease progression has been critical in the management of viral infection through ARVs and in the continued efforts to develop an efficacious preventative vaccine.

1.3.1. HIV-1 genome

The HIV-1 genome encodes a total of nine overlapping open reading frames, encoding structural (*gag*, *pol*, *env*), regulatory (*tat and rev*) and accessory (*vif, vpu, vpr and nef*) genes, which result in at least 16 different proteins (Figure 1.3). The *pol* encodes enzymatic proteins, reverse transcriptase (RT, p55) and RNaseH (p66), protease (PR, p15) and integrase (IN, p32). The *gag* (p55) is cleaved into structural proteins matrix (p17), capsid (p24), nucleocapsid (p7) and smaller proteins p6, p1 and p2 (34). The *env* (gp160) encodes the gp120 and gp41 subunits.

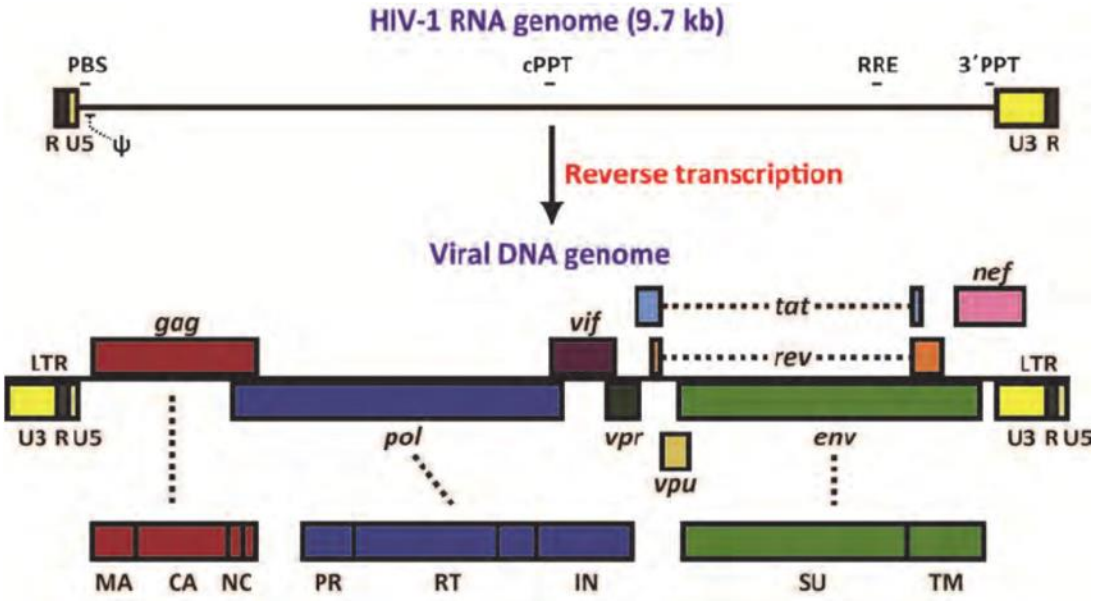

Figure1.3:Genomic organisation of HIV-1 (approximately 9.7kb), showing the resultant 14 cleavage proteins (*RnaseH* and *p6* not shown) from three main structural genes consisting of *gag, pol* and *env,* regulatory genes; *tat, rev* and accessory genes (*vif, vpr, vpu, nef)* needed for effective HIV replication. Copied from Suzuki *et al.,* 2006 (35)**.**

1.3.2. HIV-1 structure

Morphologically HIV-1 resembles most other retroviruses and is approximately 100-120nm in diameter. Two positive sense RNA single strands (approximately 9.7 kb) are surrounded by a conical shaped p24 capsid protein comprising of approximately 250 hexametric rings (Figure 1.4). The p24 also contains various viral proteins required for viral replication, such as RT and Vif. This is in turn surrounded by a matrix protein layer (MA) formed roughly of two thousand copies of the p17 structural protein (36). Enclosing the matrix structure is a lipoprotein bilayer envelope, consisting of lipids and proteins derived from the host cell membrane that are incorporated through the budding process of forming new viral particles. The incorporated host cell derived proteins include, but are not limited to, major histocompatibility complex (MHC) class I and II antigens, adhesion proteins, actin and ubiquitin. Integrated into the lipid bilayer envelope are approximately 72 viral encoded envelope glycoprotein (Env) trimers. These trimers are composed primarily of protruding surface Env trimer (SU, gp120) non-covalently bond to three transmembrane Env subunits (TM, gp41) (37). The Env is the most variable part of the virus with gp120 divided into five highly variable regions (V) interspersed by five constant

regions. The loose association between the gp120 and gp41 Env subunits results in the shedding of gp120 during the HIV replication cycle into the immediate environment.

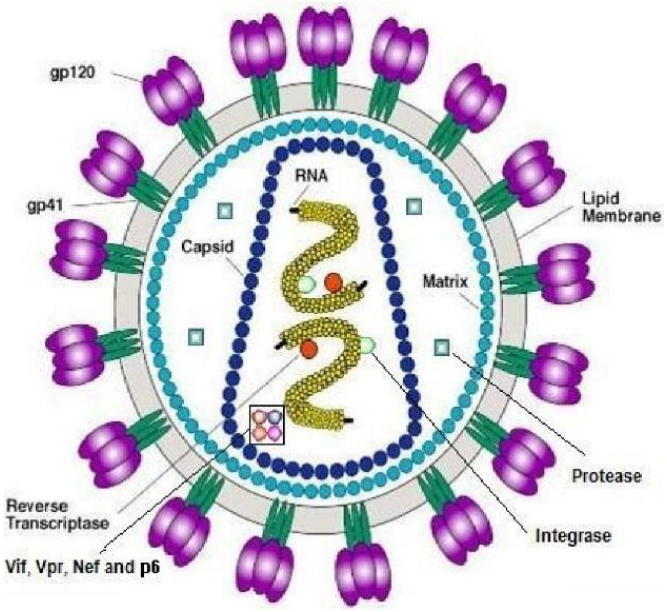

Figure 1.4: HIV-1 virion structure (approximately 100-120nm). A p24 capsid encapsulates two single stranded RNA molecules along with some enzymes required for reverse transcription during the replication cycle (RT, Vif) and subsequent cDNA integration (Integrase, p6). The capsid is surrounded by a matrix (p17) which is in turn surrounded by a lipoprotein bilayer envelope to which spikes of gp41 attached to a gp120 subunit are anchored. Copied from Mandel *et al.,* 1997 (38).

1.4.HIV-1 replication cycle

HIV-1 primarily infects CD4+ T cells. The viral life cycle is initiated by viral entry followed by reverse transcription of viral RNA to double stranded cDNA, irreversible integration of viral cDNA into the host chromosome, transcription and translation of the viral genes into proteins, assembly and ultimately budding and maturation (Figure 1.5). The life cycle takes approximately 48-72 hours, with up to 10 billion particles being produced.

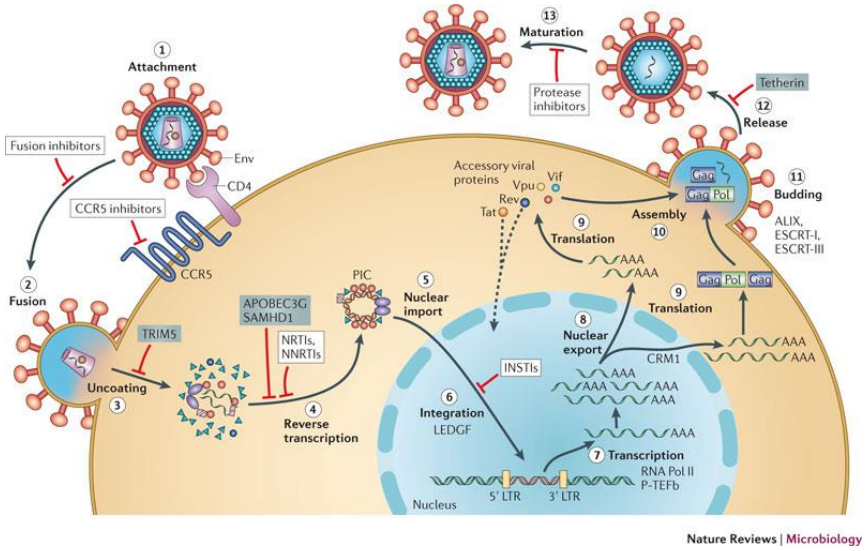

Figure 1.5: Replication cycle of HIV-1 showing entry, reverse transcription, integration, transcription, translation, assembly and budding. In the first stage of the HIV-1 replication cycle the gp120 subunit sequentially binds to the host cell CD4 receptor and the chemokine receptor for viral entry. Following gp41 mediated membrane fusion, the p24 is inserted into the cytoplasm, and the virus genome is then uncoated and reverse transcription to cDNA occurs. The cDNA assembles into a pre-integration complex that is bound to LEDGEF which aids in the transport into the nucleus where irreversible integration of the viral DNA is achieved. The viral mRNA is transcribed and transported out of the nucleus where proteins are translated. The proteins assemble at the cell membrane to form new virion particles and bud from the surface of the host cell, releasing immature virion particles. Copied from Laskey SB and Siliciano RF *et al.,*2014 (39).

1.4.1. Viral entry

HIV viral entry serves as the first stage of the viral replication cycle and can be divided into three sequential steps. First the virus interacts with cellular attachment factors, DC-SIGN and α4β7 (40, 41) of T- lymphocytes and macrophages, which are primary targets of the virus (42). The interaction brings the viral envelope in close proximity to cellular receptors, thus enhancing the efficiency of subsequent receptor recognition and binding.

Secondly, viral envelope gp120 trimer binds to CD4 receptors located on the membrane surface of the cells (43). Binding of the gp120 subunit to the CD4 results in conformational changes in the variable loops, V1/V2 and V3 of gp120 (44, 45). All these

changes allow for co-receptor recognition and the formation of an envelope flexible state required for co-receptor binding (46). The gp120 binds to chemokine co-receptors, either CCR5 (Chemokine c-c motif receptor 5) or CXCR4 (CXC chemokine receptor 4) (47-49). Binding to co-receptors initiates additional conformational changes that expose the hydrophobic pocket of gp41, creating a six helix bundle, and allowing for viral-cell fusion (50-52).

1.4.2. Uncoating and Reverse Transcription

Following membrane fusion, the p24 capsid which houses the viral RNA and other relevant viral proteins are released into the host cytoplasm. In an uncoating process, which is still not well understood, the viral core then releases the viral RNA and associated viral proteins in preparation for reverse transcription of the viral genome (53, 54). RT, which is responsible for transcribing the viral single strand RNA genome to a complementary linear double stranded DNA, has two important enzymatic components: (i) DNA polymerase, which makes copies of the viral RNA genome functioning as the synthesis template and (ii) RNaseH which cleaves the RNA template after completion of transcription (55).

The minus strand is synthesised first following initiation at the 18-nucleotide primer tRNAlys.3 which is complementary to a primer binding site (pbs) located near the 5' end on the RNA template (56). DNA synthesis extends from the pbs to the end of the viral genome, where it is then translocated to the 3' end of the same or second strand through a bridge created by the repeat sequence region, 'R' (57). Following completion of this synthesis, all RNA-DNA complexes are degraded by RNaseH apart from the polypurine tract (PPT) located at the 3' end of the minus strand template (Figure 1.3). The PPT serves as a primer for the synthesis of the positive strand to completing synthesis of the complementary linear double strand DNA genome (cDNA) (58). After completion of cDNA synthesis, the viral cDNA assembles with a range of host and viral proteins to form the pre-integration complex (PIC).

HIV-1 RT is error prone and is associated with error/mutation rates of 2 X 10^{-3} to 10^{-5} per nucleotide copied/replication cycle (59). The RT DNA polymerase like all other RNA polymerases lacks 3'-5' exonucleolytic proofreading ability which serves as a quality control parameter during DNA synthesis, aimed at repairing any replication errors. The

inability to recognise and correct replication errors during strand synthesis contributes to high probabilities for incorrect base 'reading' and subsequent nucleotide additions, which all result is multiple variants or quasispecies in an infected individual (60).

1.4.3. Integration

In the cytoplasm, a stable complex is formed between integrase (IN) and a DNA sequence within the LTR region of the newly synthesised viral cDNA (61-63). Following stable complex formation, in a process catalysed by viral IN, the 3' ends of the viral DNA undergo endo-nucleolytic processing to remove dinucleotides near a highly conserved CA region exposing the hydroxyl-OH group, in a process termed 3' end processing (64). The cDNA is associated with a PIC comprising of viral and host proteins. The exact composition of the PIC in still unclear but it houses all proteins essential for cDNA integration such as IN and Vpr (65-67), and the host cofactor Lens Epithelial Derived Growth Factor (LEDGF)/p75 which tethers HIV-1 IN to chromosomal DNA (68). The PIC contains nuclear localization signals and plays a critical role in transfer the viral cDNA from the cytoplasm into the cell nucleus in an energy dependent trans-esterification reaction. The viral DNA is irreversibly integrated into the host cell DNA genome by a process catalysed by multimeric viral IN to generate proviral DNA. Integration is critical to preventing viral genome degradation by nucleases (69, 70).
Once integrated into the host genome, the provirus remains dormant, until the cell becomes activated by the presence of transcription factors.

1.4.4. Transcription and Translation

Integrated proviral DNA serves as a template for transcription mediated by promoters in the 5' LTR region, producing a range of spliced, partially spliced and unspliced mRNA transcripts. Early protein transcripts of Tat, Rev and Nef are fully spliced, while late structural and enzymatic viral proteins are generally singly spliced or remain unspliced. Tat is one of the first proteins that is translated, which then re-enters the nucleus and binds to the TAR element in the U5 region of the LTR and significantly increases the transcription of viral RNA by recruiting positive transcription elongation factor b (PTEF-b) to the TAR element. This induces phosphorylation of residues within RNA polymerase II, stimulating elongation and increased transcription. As the concentration of Rev increases above a specific threshold it binds mRNA to transport unspliced and partially

spliced RNA from the nucleus to the cytoplasm in a Rev dependent pathway route. Completely spliced mRNA on the other hand is exported from the nucleus using normal export routes.

Viral proteins are translated in the cytoplasm, and some proteins undergo post translational modifications such as glycosylation and myristolation. Structural proteins Gag, Pol and Env are produced from the full-length mRNA. The Env polyprotein is transported through the endoplasmic reticulum to the Golgi complex where it is cleaved by cellular proteases such as furin and processed into the two Env subunits, gp120 and gp41. The viral protease initiates proteolytic cleavage of the Gag-Pol polyprotein.

1.4.5. Assembly and Maturation

Following the relevant proteolytic cleavage by viral or host proteases, the newly formed viral proteins migrate to the host membrane for assembly into virions. The gp41 anchors gp120 into the membrane of the infected cell. During assembly, viral and cellular components are packaged at nucleation sites leading to the formation of immature virions. As the new immature virus buds from the cell, it takes with it the viral Env trimers firmly embedded in the host membrane. During the maturation process, the immature virion undergoes dramatic morphological changes to become infectious. Maturation is completed after budding by the full cleavage of HIV-1 Gag and Gag-Pro-Pol polyproteins by the viral protease. The mature virion is then able to infect another cell.

1.4.6. Consequences of viral replication

A key consequence of the viral replication cycle leads to extensive viral evolution and diversity, which allows HIV-1 to escape host immune surveillance, enables the rapid establishment of drug-resistant variants and presents the greatest challenge to the development of effective ARV drugs and HIV vaccines.

During transmission, productive HIV-1 infection is generally initiated by a single viral variant (71) but within weeks following infection, plasma circulating virus constitutes a swarm of highly diverse, genetically related viral populations, referred to as quasispecies (72-74). This characteristic genetic heterogeneity is inherently due to the error prone nature of HIV RT, high viral replication and turnover rates and also genetic recombination (75). Recombination occurs following co-infection of the host with at least

two genetically diverse viruses. In a template switching mechanism during minus strand synthesis, RT DNA polymerase uses portions of both RNAs as templates. The result is a recombinant genome of chimeric DNA genetically distinct from its parental genome. The frequency of HIV recombination is reportedly the highest compared to all other retroviruses and has given rise to URFs and CRFs.
The resultant HIV-1 variants have altered replication capacity and fitness (the capacity of a virus to produce infectious progeny given specific environmental pressures) relative to the wild-type virus.

1.5. Disease Progression of HIV-1 infection to AIDS

The clinical course of HIV-1 infection proceeds along defined stages to AIDS (Figure 1.6). These stages which include i) primary/acute infection, ii) chronic infection/asymptomatic and iii) advanced/symptomatic infection and AIDS, are detailed below.

1.5.1. Acute infection

Acute infection lasts 2-4 weeks or up to 6 months and is marked by peak viral loads, innate immune responses and the emergence of adaptive immune responses, including CD4+ T cell, CD8+ cytotoxic T lymphocyte (CTL) responses and HIV-1 specific antibody production (seroconversion). Diagnosis of the acute phase has proven difficult, due to the non-specific nature of the symptoms ranging from fever, a characteristic rash, diarrhoea, night sweats and others and it is for this reason that misdiagnosis is common (76).

Following HIV-1 exposure and transmission, viral replication is isolated in the mucosa, sub-mucosa and draining lymphoreticular tissues but is generally undetectable in the blood plasma. This period lasting about 7-21 days (10 days) and is termed the ‘eclipse phase’ (77) and ends following migration of the virus to gut associated lymphoid tissue (GALT) and other Lymph node areas such as the throat (78). GALT is considered the primary site of HIV-1 infection, allowing for large depletion of the CD4+ memory T cells and an exponential rise in plasma viremia, to 1-2.5 million RNA copies per millilitre (copies/ml) , as illustrated in Figure 1.6 (79, 80). During the early acute phase, a latent HIV-1 reservoir is also established (80).

Weeks following initial infection, HIV-1 specific immune responses can be detected, including CD8+ CTLs as well as binding and neutralising antibodies (81-83). The presence of HIV-specific CD8+ CTL response is responsible for controlling viral replication and decreasing viral RNA concentrations to establish a patient specific viral set-point (76).

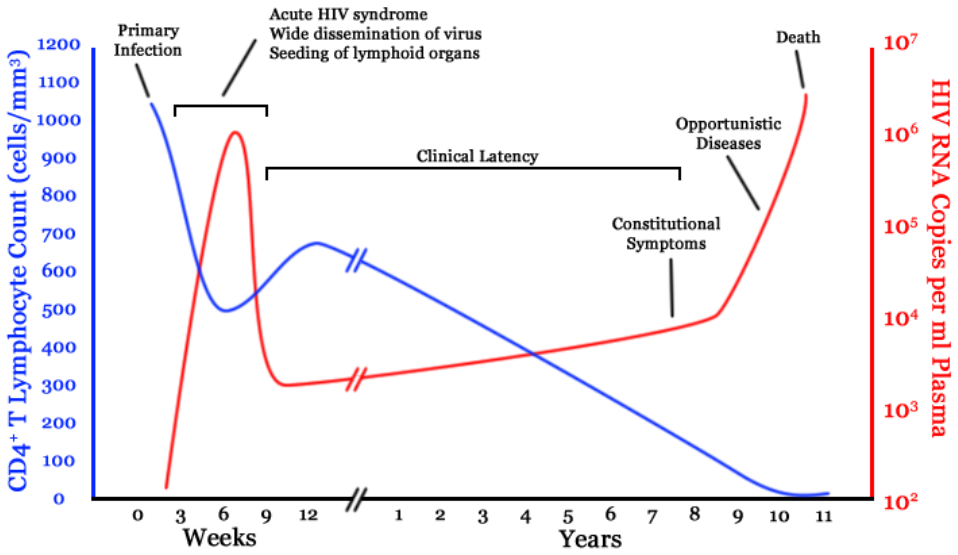

Figure1.6: The relationship between plasma circulating HIV-1 RNA concentrations (RNA copies/ml; right- Y-axis) and CD4+T lymphocyte cell counts (cells/mm^3; left Y-axis) as markers of HIV disease progression to AIDS. Copied from Naif *et al*., 2013 (84).

1.5.2. Chronic Infection/Asymptomatic phase

This phase of infection can last up to 10 years with the majority of infected individuals presenting with no clinical symptoms. The length of time individuals remain in this asymptomatic phase is highly variable and is hypothesised as a balance between virion cell infection and clearance by host immune responses (85). During this phase there is on-going viral replication (measured as viral load, RNA copies/ml) and turnover of CD4+ T cells (measured as cells/mm^3). Eventually the immune system loses the battle against HIV-1, and the CD4+ T cell counts decline to levels where the individual starts becoming symptomatic.

The specific viral set-point is considered a determinant of disease progression to AIDS (86). Higher viral set-points are linked with rapid disease progression, with individuals

developing AIDS within 3 years (86). This group of rapid progressors is made up of less than 5% of the total HIV population. Typical progressors generally develop AIDS within 10 years of infection. By contrast, long term non-progressors develop AIDS slower, >10 years after seroconversion. Additionally, Elite controllers form part of a rare group of individuals (1/400 infected individual), able to maintain high CD4 + T cell counts and low viral loads (<50 RNA copies/ml) and do not progress immunologically to AIDS. Elite controllers and long term non-progressors make up less than 5% of the total infected HIV population.

1.5.3. Advanced/Symptomatic Infection

A continued decline in the immune system eventually leads to opportunistic infections (OIs) such as candidiasis, tuberculosis, and cancers such as Kaposi's Sarcoma (87). These AIDS defining illnesses result in an AIDS diagnosis (full list available on (https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5710a2.htm). In the absence of HAART, an individual will die of AIDS within two years.

1.6. Antiretroviral therapy

Following the introduction of HAART (a three or more drug combination of two ARV drug classes), in 1996 for the management of HIV-1 infection, reduced HIV-1 related mortality and morbidity has been documented (88).

1.6.1. Currently available ARVs

ARVs target several critical steps in the HIV-1 replication cycle, thus suppressing viral replication and consequently lowering viral load. The multiple steps targeted by ARVs as shown in Figure 1.5 include, viral entry, reverse transcription, integration and protein cleavage by PR. Currently approved ARVs are classified under four drug classes: RT inhibitors (nucleoside/nucleotide RT inhibitors (NRTIs) and non-nucleoside RT inhibitors (NNRTIs), PR inhibitors, IN inhibitors, and Entry inhibitors (Fusion inhibitors and co-receptor antagonists) (Table.1.1).

Table1.1: Currently available FDA (FDA) approved HIV-1 antiretroviral drug classes, targeting different HIV enzymes and stages of the viral replication cycle**.**

Inhibitor class		**Drug Name**	**FDA approval (Year)**
RT inhibitors	**non-Nucleoside Reverse transcriptase inhibitors (nNRTI)**	Delavirdine (DLV) Efavirenz (EFV) Etravirine (ETR) Nevirapine (NVP) Rilpivirine (RVP) Doravirine	May 2001 September 1998 January 2008 June 1996 May 2011 August 2018
	Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTI)	Tenofovir (TDF) Emtricitabine (FTC) Abacavir (ABC) Didanosine (DDI) Lamivudine (3TC) Stavudine (d4T) Zidovudine (AZT/ZDV)	October 2001 July 2003 December 1998 October 1991 November 1995 June 1994 March 1987
Protease Inhibitors (PI)		Atazanavir (ATV) Darunavir (DRV) Fosamprenavir (FPV) Indinavir (IDV) Lopinavir (LPV) Nelfinavir (NFV) Ritonavir (RTV) Saquinavir (SQV) Tipranavir (TPV)	June 2003 June 2006 October 2003 March 1996 September 2000 March 1997 March 1996 November 1997 June 2015
Integrase-Strand Transfer Inhibitor (INSTI)		Dolutegravir (DTG) Elvitegravir (EVG) Raltegravir (RAL) Bictegravir (BIC)	August 2013 September 2014 October 2007 February 2018
Entry Inhibitors	**Fusion Inhibitors**	Enfuvirtide (ENF)	March 2003
	Co-receptor antagonsist	Maraviroc (MVC)	August 2007

1.6.2. Antiretroviral drug regimens

The current WHO guidelines recommend prescription of ART to any person testing positive for HIV, regardless of WHO stage or CD4 cell count (89). This 'test and treat' policy is in response to mounting evidence from both randomised observational studies showing a significantly higher reduction not only in the rate of HIV-1 sexual transmission (90) but also in morbidity and mortality associated with treatment naïve patients initiated on treatment with higher CD4 cell counts (91, 92).

The current recommended guidelines are an improvement to prior guidelines and have been adapted in response to a better understanding of HIV progression outcomes and the development of more potent ARVs. In 2010 adults were prescribed ART if presenting with a CD4 cell count of ≤ 200 cells/mm^3, while in 2013 adult patients were started on ART, when CD4 cell count was ≤ 350 cells/mm^3 (93). In 2013, WHO guidelines recommended a standardized first-line regimen of 2 NRTIs (Tenofovir (TDF) or AZT if renal failure is associated with TDF and Lamivudine (3TC)) prescribed together with NNRTI Efavirenz (EFV) or Nevirapine (NVP) if there is intolerance to EFV. Patients who fail on first- line regimen were prescribed a second- line PI regimen consisting of 2 NRTIs (AZT/TDF) and a PI Lopinavir (LPV/r) (93).

The current 2018 adult ART regimen WHO guidelines recommend a regimen consisting of 2 NRTIs (Tenofovir (TDF) and Lamivudine (3TC) given together with an integrase inhibitor, such as Dolutegravir (DTG) available already in some countries as a generic fixed dose combination (FDC). Alternatively, patients can be prescribed a first-line regimen of 2 NRTIs, TDF and 3TC/FTC with nNRTI, Efavirenz, as an FDC taken once daily (94). There is however major caution for DTG use in Pregnant women as earlier evidence points to early abortions, still births and fetal abnormalities which include neural tube defects if the drug is prescribed to women during conception or early stages of pregnancy (first and second trimester) (95). Patients who fail on first-line ART regimen are prescribed a second-line PI based regimen. WHO recommends a PI based regimen of ritonavir boosted Lopinavir (LPV/r) or Atazanavir (ATV/r) combined with 2 NRTIs for patients previously on DTG as their first-line regimen. For patients previously on a failed 2NRTI and EFV first- line regimen, a 2 NRTI and DTG regimen is recommended. The choice of NRTI depends on the previous first line regimen. If failure was on TDF then

AZT should be used and vice versa. Following second-line regimen failure, a third- line regimen of DRV/r-DTG and 1-2 NRTIs is recommended (94).

In South Africa, a standardised first- line regimen of ABC/TDF–FTC/3TC- EFV is currently being offered in the public sector. It is worth mentioning that a fixed dose of TDF-3TC-DTG (TLD) has recently been approved in South Africa and roll-out into the public sector is anticipated. AZT-3TC- LPV/r /ATV/r is prescribed as part of a PI based second-line regimen. If the patient is HBsAg positive TDF is added to the regimen as well. For patients failing on a d4T/AZT-based first-line regimen TDF is prescribed instead (96). A specialist committee is responsible for determining a patient specific third-line regimen. Selection is based on several factors including, available drugs and patient ARV drug resistance profile (based on genotyping) (97).Once initiated, an individual need to consistently take ARV drugs for life. If ARV treatment is successful, the viral loads become undetectable (generally < 20 RNA copies/ml), and CD4 T cell counts recover (but never to normal levels) within 12 weeks. However, if treatment is interrupted for any reason, the viral load will rebound within weeks (98). Unfortunately, the current ARV drug regimens cannot target the HIV-1 reservoirs, so no HIV cure is available.

1.7. Mechanism of action of antiretroviral drugs

The focus of this dissertation is on RT and PR inhibitors thus their mechanisms of actions are discussed in more detail.

1.7.1. Reverse transcriptase inhibitors (RT Inhibitors)
Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTI)

NRTIs are dideoxynucleotides (ddNTP) lacking a hydroxyl group (OH) on their sugar moiety and function as nucleoside analogues. They serve as competitive inhibitors to cellular endogenous deoxy-nucleotide triphosphates (dNTPs) which are RT substrates.

This class of inhibitors are prescribed as prodrugs (inactivated form) to increase their bioavailability. NRTIs require activation to their respective 5’- trisphosphate by host kinases and phosphorylases (99-101) through intracellular phosphorylation. Other drugs in the group, Abacavir and Didanosinerequire further changes at their base for full activation (102). The triphosphorylated form of the inhibitors acts as a chain terminator in

viral reverse transcription, therefore inhibiting proviral DNA synthesis. During the polymerization reaction by RT, dNTPs are incorporated to the growing viral DNA strand which results by formation of a phosphodiester bond between the 3' template and the 5' of the incoming nucleoside. This is followed by release of the pyrophosphate from the RT p66 subdomain active site to allow for the next dNTP to be added in the chain (103, 104). Since NRTIs structurally mimic natural nucleosides they can be incorporated into the growing strand by cellular polymerase. Their incorporation to the primer-template strand prevents the formation of a 3'-5' phosphodiester bond and the resultant chain elongation (101, 105-107). NRTI drugs have shown limited specificity for HIV-1 RT. In their activated form they not only have an inhibitory affinity for HIV-1 RT but also mitochondrial DNA polymerase and HIV-2 RT (108-112). Inhibition of mitochondrial DNA polymerase is a major problem with a cascade of adverse effects, leading to high level drug toxicity and cell death (112).

Non -Nucleoside Reverse Transcriptase Inhibitors (nNRTI)

nNRTIs belong to an array of structural classes from different chemical families (113-115). The group may show chemical heterogeneity, but all members bind allosterically to a specific hydrophobic pocket in the p66 subdomain, ~ 10A° (develops only in inhibitor presence) from the polymerase catalytic site of RT. For this reason, they serve as non-competitive inhibitors to RT catalytic activity. It is known that binding of the drug compounds to the pocket induces several conformational changes to the active site. Each of these changes has been hypothesised as a mechanism of inhibition to RT catalytic activity. Kohlstaedt *et al.,* (1992) suggested that binding of the drugs to the hydrophobic pocket, which when present serves as a hinge between the RT palm and thumb subunits (p51/p66) decreases thumb mobility (116). This restricted mobility slows or even possibly prevents primer/ template translocation and elongation of the strand. Hsiou *et al.,* (1996) proposed that binding causes distortions to the precious geometry of the polymerase active site and displacement of the p66 and p51 domains. These distortions slow DNA polymerase activity and decrease RNaseH cleavage specificity (117).

Despite the mechanism proposed for the inhibition it is clear that binding of the drug affects the rate-limiting step in the catalysis reaction during the steady state (106, 118-121). The hydrophobic pocket has been shown to exist only in the presence of nNRTIs

thus making them exclusively active against HIV-1 (113, 122, 123). Also, unlike NRTIs they do not undergo intracellular phosphorylation for activation of their inhibitory properties.

1.7.2. Protease Inhibitors (PI)

PIs are highly selective, competitive inhibitors, mimicking the tetrahedral transition state structure of the HIV-1 PR natural substrates, competing with them for binding to the active site of the enzyme. The FDA-approved HIV protease inhibitors share same structural similarities and a similar binding pattern. PIs cannot be cleaved by PR and thus block enzyme activity for the cleavage of Gag and Gag-Pol polyprotein precursors crucial for viral assembly and maturation for infectious virions (124).

Currently available PIs are designed with a boosting agent, ritonavir. The pharmacokinetic property of ritonavir is to block the rapid metabolism of PIs by the cytochrome P450 3A4 (CYP3A4; enzyme involved in drug and toxin metabolism in the hepatic system) and allows for better absorption of these PIs. Lower less frequent doses can thus be prescribed to patients with the expectance of better adherence and minimal toxic effects. The most commonly used PIs are LPV/r and ATV/r.

1.8. Drug resistance

A major limitation to sustained HAART regimen efficiency is the inevitable emergence of HIV-1 drug resistant mutations (DRM). HAART failure is defined in low-middle income countries (LMICs) as two successive viral loads ≥ 1000 RNA copies/ml, and is attributed to associated drug toxicity, poor viral response or patient non-adherence resulting in sub-optimal drug concentrations. An updated list of mutations that reduce the level of susceptibility to each of the currently approved FDA ARV drugs is available (125) and can be accessed online (https://hivdb.stanford.edu/).
An HIV-1 infected individual can either acquire DRM following HAART initiation (acquired DR), or an uninfected individual can become infected with drug resistant virus (transmitted DR, TDR).

Following the introduction of HAART, selective drug pressure exerted by each inhibitor mandates a shift in the quasispecies equilibrium. Viral populations with mutations that

confer resistance to the viral inhibitors are primarily selected and may emerge to compete for dominance in the new ARV drug environment. Incomplete suppression of viral replication subsequently allows for increased frequency of the mutant populations that result in reduced drug susceptibility while maintaining high viral fitness (ability to replicate) and inevitably led to treatment failure.

Viral drug resistant strains from these treatment experienced patients can be transmitted to uninfected individuals, in which they persist as either minority or dominant viral populations. The TDR variants may revert to the wild-type virus, at a rate dependent on their cost on viral fitness. NRTI TDR mutations (TDRM) such as M184IV have been shown to revert to wild-type virus in less than 1 year while nNRTI and PI mutation can take up to 3-10 years (126). nNRTI TDRMs are the most common of all drug class TDRMs (127).

The emergence of TDRMs in treatment naïve patients impacts on future ART success, as it has been shown to lead to early treatment failure in these patients, thus limiting ART regimen options for such patients (128). Also, an increase in the frequency of TDR amongst a country's HIV infected population maybe be evidence of a poor ART programme, with factors such as a patient retention, poor ART patient adherence and inadequate resistance monitoring. WHO recommends a change in a country's first-line ART regimen if the prevalence of TDR is ≥ 10% of the HIV infected population (129).
DRMs exerting intermediate to high-level DR are classified as primary DRMs and not only reduce drug susceptibility but also negatively impact on viral fitness. Given the negative effects on target viral fitness, compensatory/ secondary DRMs develop in an effort to restore viral fitness to levels similar to wild type virus or in some case even higher.

The extent to which the accumulation of DRMs decrease drug susceptibility and eventually lead to ART regimen failure, mainly depends on the genetic barrier of the inhibitor, as some inhibitors require more than one DRM to confer significant drug resistance.

1.8.1. Resistance to Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTIs)

Drug resistance by NRTIs is achieved by two main mechanisms: i) an Increased discrimination between natural dNTP substrate and inhibitor analogues by RT, and ii) pyrophosphoric excision of the NRTI-MP (monophosphate) from the chain terminated primer. As mentioned previously, NRTIs are competitive analogues to the RT natural substrate, inhibiting chain elongation of cDNA synthesis. Mutations mostly located in and around the DNA polymerase active site of RT (Including M184V, K65R, Q151M and L74V, Figure 1.7) facilitate the preferential discrimination of NRTI-MP over a dNTP in analogue binding during the reverse transcription reaction

The second mechanism of resistance selected for by thymidine analogue drugs (such as AZT and d4T) increases the rate of nucleoside analogue excision. The group of mutations involved in the mechanism includes mutations at positions, 41, 67, 70, 215, 210 and 219 better known as thymidine analogue mutations (TAMS, Figure 1.7). The presence of TAMs increases the rate of pyrophosphorolytic excision which is an ATP/PPi mediated ddNTP excision by HIV RT. ATP/PPi acts as a nucleophile, attacking the chain terminated primer/template in a pyrophosphorolytic excision reaction, forming a dinucleotide triphosphate and tetraphosphate, respectively. This excision allows for continued chain elongation as the primer is free to incorporate an incoming dNTP (130, 131).

1.8.2. Resistance to non-Nucleoside Reverse Transcriptase Inhibitors (nNRTIs)

nNRTIs inhibit reverse transcription by binding to a hydrophobic pocket near the active site of HIV RT and reducing its mobility for substrate binding. Mutations in this drug class include, K101ERQ, V106A and E138AG35 (Figure 1.7) and prevent the formation of the hydrophobic pocket which is only visible in inhibitor presence, thus resulting in a functional RT enzyme for reverse transcription (60, 132). Another nNRTI mutation mechanism for reducing drug susceptibility is attributed to the emergence of K103N. This highly prevalent nNRTI mutation at position 103 is located near the entrance of the hydrophobic pocket where it creates a hydrogen bond with the unliganded RT. This blocks nNRTI access to the pocket and thus the inhibitor fails to reduce enzyme efficiency (60).

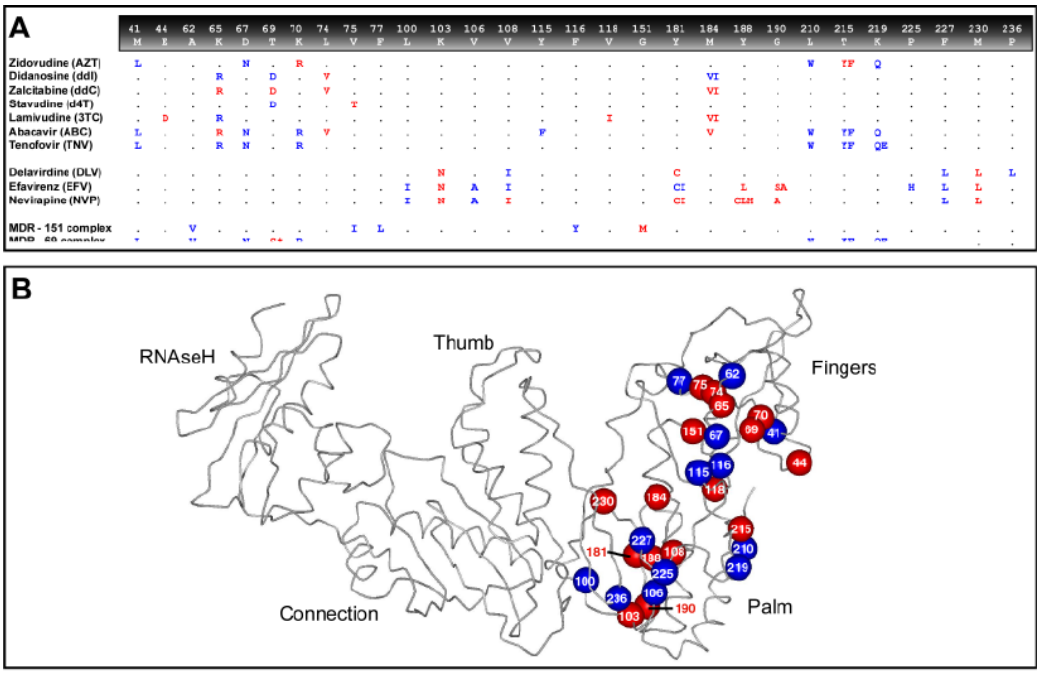

Figure1.7: HIV-1 RT enzyme showing location of primary (highlighted in red) and secondary (highlighted in blue) RT nNRTIand NRTI mutation residues.**A**. lists the RT mutations to selected RT inhibitors. **B**. shows the location of each mutation residue on RT protein. Copied from Quiñones-Mateu *et al*., 2001 (133).

1.8.3. Protease Inhibitors mutations

More than 20 amino acid substitutions at the 99 amino acids of HIV-1 PI have been associated with PI drug resistance. This includes both primary and secondary/compensatory PI mutations as shown in Figure 1.8.

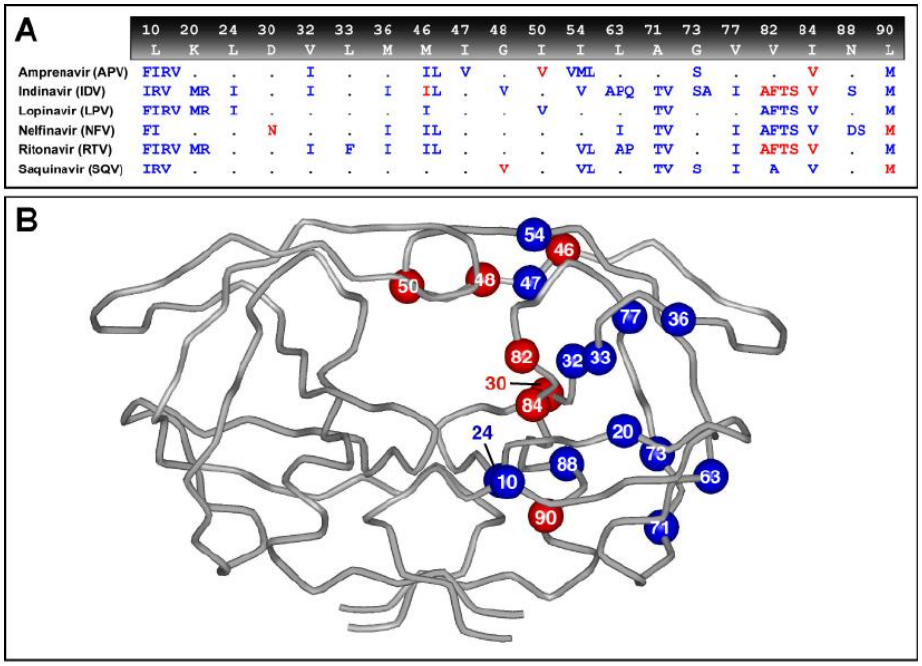

Figure 1.8: HIV-1 protease enzyme showing location of primary (highlighted in red) and secondary (highlighted in blue) PI mutation residues. **A**. lists the PR mutations to selected PR inhibitors. **B**. shows the location of each mutation residue on PR protein. Copied from Quiñones-Mateu *et al.,* 2001 (133).

Most primary PI DRMs are located at the PR active site (residues 25–32, 47–53 and 80–84) and around the active site (for example, G148V and I54V are located in the flap region of PR) at PR inhibitor binding residues resulting in inefficient inhibitor binding. These primary residues have been shown to be sites where inhibitors protrude beyond the substrate envelope and conformational changes mostly affect binding and recognition of the inhibitor while still able to recognise the natural substrate Gag and Gag-Pro-Pol polyprotein (134-136).

A study by Prabu-Jeyabalan *et al*., (2003), showed the crystal structures of three natural substrates and inhibitors SQV and RTV in complex with an inactive drug-resistant mutant (D25N/V82A) HIV-1 protease (137). They noticed that Val82 residue reduced the van der Waal interactions between Inhibitor and enzyme and resulted in conformational changes in the Cα backbone compared to the wild-type structure. All the changes impacted largely on inhibitor binding compared to the natural substrate peptides as inhibitor could not fit in the active site.

In another study by Schiffer *et al.,* (2010), the binding of inhibitors, FPV, DRV, ATV and SQV to the PR variants G48V and I54V located in the flap region of PR caused conformational changes that locked the flaps in an open configuration.

The development and combination of some DRMs are considered signature profiles to each PI. For example, DRMs, I50V and I50L are selected and found frequently in patients failing FPV/DRV and ATV therapy, respectively, while mutations at PR residue 30 are observed in patients failing NFV.

Most PI DRMs confer resistance to more than one PI resulting in cross-resistance to varying degrees as most of the inhibitors bind to similar PR residues. *In vitro* susceptibility testing on more than 2,400 isolates for susceptibility to FPV, IDV, NFV, LPV, ATV and DRV was performed by Rhee et al., (2010). They reported on 46 known mutations at 26 positions of PR that conferred resistance to one or more PIs. They showed that mutation I84AV had the highest cross-resistance, with decreased susceptibility to eight PIs. Mutations V32I, G48V, I54ALMSTV, V82F and L90M were all associated with decreased susceptibility to six to seven PIs mutations while PI mutations D30N, I50L and V82AL showed the least level of cross-resistance, as they were associated with decreased susceptibility to fewer than four PIs (138).

The extent of resistance conferred by each primary mutation is different for each PI, with some conferring high-level resistance and others low-level resistance. Mutation at residue 46, confers potential low-level resistance to LPV/r and ATV/r; mutation V82A confers intermediate resistance to LPV/r but low-level resistance to ATV/r, while mutation I84V confers high-level resistance to all PIs.

The emergence of PI DRMs not only reduces PI antiviral activity but has a huge cost on viral fitness. Secondary mutations in the PR emerge to compensate for HIV-1 replication and maturation impairment are generally located outside the substrate-binding cavity of the enzyme (139-141). It has been proposed that secondary mutations alter the active site of PR in order to adapt to the changes introduced by primary mutations. Clemente *et al.,* (2003)engineered HIV protease mutants containing PI DRM, D30N with and without M36I and/or A71V (142). The K_i values (measure of binding affinity of inhibitor for the enzyme) for the inhibitors NPV, RTV and IDV as well as the catalytic efficiency of the mutants was determined. They found that HIV variants containing M36I or A71V alone did not significantly change in the binding affinities to any of the inhibitors tested. However, the D30N mutant had reduced inhibitor binding affinity. The double mutants

containing a combination of mutations D30N-M36I, and D30N-A71V displayed increased binding affinity, thus highlighting the compensatory effect of A71V and M36I.
Other compensatory mutations apart from M36I and A71V that have been identified include, M46I and V82T. Compensatory PI mutations increase viral enzyme efficiency but also confer PI drug resistance but not significant enough on their own, requiring the development of primary PI resistance in order to achieve intermediate- to high-level resistance to the inhibitors.

The accumulation and appearance of PI mutations is in a step-wise ordered manner (143-146) of both primary and secondary mutations, where each inhibitor selects for signature patterns of mutations. Mutation V82A has been shown to develop first against RTV, with Molla *et al.,*1996(143) showing its development as early as 3 weeks in patients receiving suboptimum concentrations of RTV (500mg). This was followed by the appearance of I54V, M36I and A71V by week 8 with the latter two considered as compensatory mutations.

Drug resistance to PIs requires more than four primary drug resistant mutations and secondary mutations to confer intermediate- to high-level resistance while maintaining viral replication capacity and maturation (138, 143, 147-151). Mammano *et al.,* (2000) examined the effect of single and combined amino acid substitutions in HIV-1 PR on both resistance to PIs and on drug-free infectivity. The presence of V82A conferred a slight increase in resistance to RTV compared to wild-type virus. V82A in combination with two or more mutations (I54V, A71V, M46I) was required for significant resistance to RTV (141). The high genetic barrier characteristic of PIs and the impact on protease catalytic activity, could explain the uncommon frequency of PI resistance in patients more especially at low-level adherence.

PI based regimens are highly potent and evidence has shown that most patients virologically failing on the regimen have no PR DRMs conferring PI resistance and that non-adherence in most patients explains the observed virological failure (152-159). El-Khatid*et al*., (2010) conducted a cross-sectional study among patients receiving ART for ≥12 months and genotypic resistance testing was performed on individuals with a viral load > 400 RNA copies/ml. For patients virologically failing a second-line PI based regimen (34/115), nNRTI and NRTI mutations occurred at a prevalence of 29% and 54%, respectively, while PI mutations were observed in only 6% of the patients. They

concluded that non-adherence may be the main reason for the observed virological failure. In a different study by Van Zyl *et al*., (2011), they demonstrated that most patients experiencing PI- based regimen virological failure had low LPV/r plasma and hair concentrations, thus non-adherence was a major cause of failure in these patients. Of the 93 patients included in their study, 37 experienced virological failure, and only 2 harboured PI DRM. LPV/r Plasma concentration was available for one of the two patients and was 10.3 µg/mL. They predicted that virological failure was associated with low LPV/r concentrations of ≤1 mg/mL or hair concentrations ≤3.63 ng/mg at a viral load cut off of 1000 RNA copies/ml (153).

1.8.3.1. Gag-Pol cleavage site drug resistance mutations

It has been previously shown that DRMs in the cleavage site of Gag-Pol emerge to compensate for a loss in viral fitness on account of DRMs in PR. Evidence now also demonstrates that some of these mutations also confer DR to PIs (160-166). It is proposed that mutations in the Gag-Pol cleavage site indirectly improve the ability of the mutated PR to preferentially bind the natural substrate over the PIs (60, 141, 167, 168) . However not all Gag-Pol cleavage site mutations reduce the level of PI susceptibility while compensating for fitness loss (141). Overall, the level of resistance conferred by Gag-Pol cleavage site mutations, like most PR secondary mutations, does not significantly reduce susceptibility on their own.

1.9. Genotypic monitoring of ARV treatment failure

Clinical monitoring of patients (viral loads and CD4 T cell counts) is used to predict ARV treatment success or failure. Subsequently, genotypic resistance testing in patients failing HAART can predict the presence of DRMs and inform future regimens. Currently in the public health sector in South Africa, genotypic testing is only performed for patients failing a second-line PI based regimen and those on third-line salvage therapy. This is mainly attributed to the high assay costs, and moreover, the pattern of DRMs following first-line regimen failure is fairly predictable and has been well documented. The FDA approved gold standard genotyping assays use population- based Sanger sequencing which only detects viral variants present at ≥ 15-20% of the viral quasispecies (169-171). More sensitive technology and techniques have become available and allow for genotypic testing and detection of HIV-1 DRMs below the limit

detection of population-based Sanger sequencing which are referred to as minority HIV-1 variants.

Minority HIV-1 variants with DRMs have been detected using a variety of methods, including point mutations assays, including qPCR, allele-specific PCR, oligonucleotide ligation assays, single genome amplification and sequencing, and next generations sequencing (NGS). These assays have varying sensitivities, but most can detect variants at proportions of ≥ 1% of the quasispecies. The two most commonly used commercially available next generation sequencing platforms are shown in Table 1.2.

Table 1.2*:* **Comparison of Illumina Miseq and Ion torrent NGS instrument properties.**

Properties	**Illumina (MiSeq)**	**Ion Torrent (S5 and S6)**
Amplification method	Bridge PCR *in situ*	Emulsion PCR on beads
Principle (chemistry)	Sequencing by Synthesis (SBS) (reversible termination)	Synthesis (H+ detection)
Average read length (bp)	150 - 300	200 - 400
Primary error (error rate)	Substitution ~0.1 %	Indel ~1 %
Main advantage(s)	Easy work flow, maturity	Low cost, fast run
Main disadvantage(s)	Shortest reads	Homopolymer misreads

1.10. Illumina MiSeq Sequencing by synthesis chemistry

The Illumina MiSeq sequencing by synthesis (SBS) protocol was used in this study and consists of three steps: (i) sample/library preparation (ii) cluster generation and (iii) sequencing (172, 173).

First in a Library preparation step, input DNA is randomly fragmented to varying lengths using an enzymatic reaction utilising a transposome enzyme. Thereafter sequence adapters are attached to both ends of the DNA fragments and denatured to single strand DNA using PCR amplification. These are then indexed with unique identifiers and universal primers complementary to oligo nucleotides, densely covering the flow cell of the sequencer. AMPURE XP beads are then used to select the required size fragments while also purifying the library.

The prepared library is ligated to the flow cell and amplified using isothermal 'bridge amplification, which makes approximately 1000 clones of each fragment in a cluster generation process (174). Initial amplification of the template strand from the 3' end, attached to a p5 oligonucleotide precedes a denaturation and wash step of the original template strand with formamide. As illustrated in Figure 1.9, the 5' end of the new strand anneals to the flow cell surface, its 3' end anneals to the complementary oligonucleotide sequence at p7 on the flow cell surface, forming a bridge and serves as the template for the next synthesis. Following several amplification cycles, of ~ 32 amplification cycles, reverse terminator-based sequencing is completed. Four reversible terminator nucleotides each bound with a different fluorescent dye are incorporated into each DNA strand. A schematic illustration of clustering by bridge amplification is shown in Figure 1.9 (173, 175).

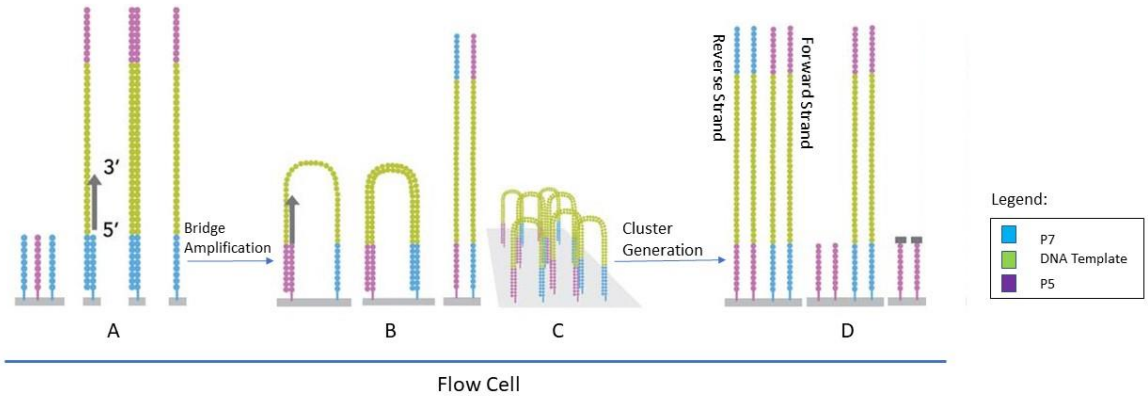

Figure 1. 9: **A schematic representation of bridge amplification**. Copied from Illumina (175).(**A**). The DNA template hybridizes to complementary adapters on the flow cell of the illumina instrument. (**B**) The DNA strand is extended in the 5'-3' direction and in a bridge amplification thep7 primer bends over to attach to the binding site on the opposite end of the fragment, and the p5 index is extended. The newly formed DNA strands are denatured, cleaved and the original strand is washed away. (**C**). The amplification cycle repeats for both DNA reads until millions of copies of each DNA template are generated. (**D**) Each of the resultant strands is to be sequenced in a Synthesis by (SBS)

1.10.1 Sequencing errors

A population-based sequencing, NGS is associated with error rates which impact on the correct calling of each nucleotide base added during the sequencing process. These error rates vary from each platform and sequencing technique. Nevertheless, the most common errors include those introduced during PCR amplification, during sample preparation and during sequencing of the clusters.

During sample preparation, viral RNA is converted to cDNA by RT-PCR amplification using high fidelity transcriptase enzymes. Despite the high fidelity of the polymerase, substitutions, insertions and deletions may be introduced during this reaction. PCR bias from primer mismatch is known to result in preferential amplification of some variants

over others (176). Again, sequencing of each cluster may result in phasing errors following nucleotide incorporation and terminator excision. A single nucleotide is incorporated per cycle to each growing strand, followed by a wash step to remove unincorporated nucleotides, in preparation of the next cycle. As illustrated in Figure 1.10, the terminator may not be completely removed after nucleotide incorporation and thus no other nucleotide can be incorporated during the next cycles. This result in a shorter sequence compared to those in the cluster and this is termed as Pre-phasing. Again, unincorporated oligonucleotides may not be completely washed away after each cycle, and thus can be incorporated into the chain, resulting in a sequence, two or more nucleotides longer than the rest in the cluster after the cycle (177) (Figure.1.10).

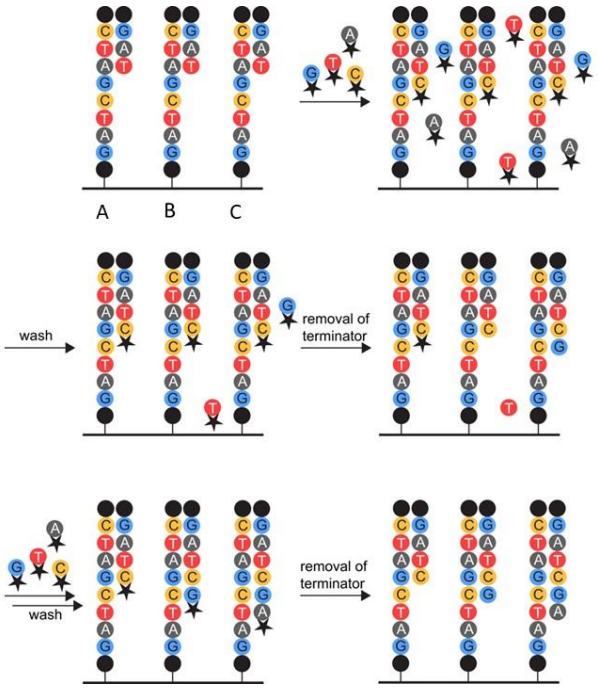

Figure 1.10: **Depiction of the sequencing-by-synthesis approach. The black dots represent the sequencing primers.** The terminator (black star) on the deoxynucleoside triphosphates (dNTPs) prevents the addition of the subsequent nucleotide to the growing DNA strand. (**A**) Depicts a post-phased sequence. The terminator is not completely removed during the wash step and no nucleotide can then be incorporated during the next sequencing cycle. This results a shorter DNA sequence. (**B**) depicts the state without phasing effects of any kind. Following nucleotide addition, the terminator is removed, and after a wash step the next nucleotide is added to the growing sequence chain. (**C**) If non-incorporated nucleotides remain following the wash step an additional nucleotide may be incorporated in a single cycle thus the resulting

strand will subsequently be longer than the original and is post-phased. Copied from Preiffer*et al.,2018* (177).

SBS chemistry is associated with a 0.1% error per nucleotide per cycle. For this reason, the quality of the sequence run is measured often by Q- scores. A Q30 score, which represents a 0.001% probability that a nucleotide was incorrectly called, is most commonly used. The higher the Q30 score the more reliable is the run data. Illumina is mostly associated with Q30 score of ≥ 80%, but scores of ≥65 have been shown to still provide useable sequence data. Most bioinformatics software's correct for the possible errors of NGS, with, for example, short sequence reads being removed from analysis and bases not complementary being discarded. A minimal mutation cut-off of 1% is often used for NGS data, as sequencing errors have been shown to increase below this percentage.

1.11. Relevance of minority HIV DR variants

Several studies have detected the presence of minority HIV variants with DRM in both treatment naïve and treatment experienced patients. Evidence from such studies has shown a strong associated between the presence of minority variants and ARV treatment regimen failure for such patients, particularly for those on nNRTI based regimens.

In the OCTANE Trial, they aimed to determine the efficacy of a nevirapine (NVP) based regimen to that of LPV/r in women infected with HIV-1, who had or had not taken single dose NVP at least 6 months before enrolment. They found significantly more women in the NVP group reached the primary end point (had virological failure or died) with 26% vs. 8% in the LPV/r group (128). In a follow up study, highly sensitive allele-specific PCR was used to detect DR mutations in 201 women who were previously reported to have no NVP resistance by Sanger sequencing. Among these women, 70 (35%) had NVP-resistant mutants detected by allele specific PCR. The occurrence of virological failure in the NVP group was significantly associated with the presence of NVP-resistant mutations K103N or Y181C at frequencies >1%. Therefore, among women with prior exposure to NVP, minority NVP-resistant mutants were associated with increased risk for failure on an NVP-containing regimen (178).

This study (179) and others (170, 179-183) provide evidence for the clinical significance of HIV minority variants in predicting HAART regimen failure. They also highlight the benefit of using ultra-sensitive NGS assays for HIV genotyping of minority variants and the selection of patient specific HAART regimens. Similar links have also been made in ART experience patients

In contrast, studies investigating a similar link in PI experienced or naive patients have failed to provide strong evidence pointing to such an association between minority PI variants and virological failure of patients on PI regimens (155, 184-187). Dykes et al., 2004 (188) evaluated the presence of minority variants in 10 patients on a PI based regimen of AZT-3TC-IDV, for whom first line AZT had failed. Baseline sequencing using bulk population sequencing prior to PI regimen initiation showed 4/10 patients had detectable PI mutations. PI mutations detected in these patients were, however, secondary mutations, M36I, A71T/V and accessory mutation K20R. Patients also had RT mutations indicative of resistance developed at first-line ART failure. At failure on the PI based regimen only four of the 10 patients had primary PI mutations detected. Of the 6 without primary resistance mutations (as previously determined by bulk sequencing), sequence clone analysis was performed and for one patient, V82A was detected. Interestingly, patients were kept on a failing IDV regimen and in follow up analysis 16 weeks later, the frequency of V82A in the patient did not increase and no other PI mutations had developed. The study failed to show an association for the increased virological failure with the detection of minority PI variants.

Similarly, Fisher et al., 2012 (155) compared the genotypic results of population based Sanger sequencing and Roche 454 ultra-deep pyrosequencing (UDP) following a PI based regimen failure in seven patients. Patients had been on a PI based regimen for a median of 16 months and had been failing the regimen for an average of 10 months. Population based Sanger sequencing performed at first-line failure revealed the presence of expected RT associated mutations (M184V/I and TAMs) and no major PI DRM. At second-line failure, population-based Sanger sequencing detected no PI DRMs. However, UDP detected DRMs, I54T, M46V, I54T, V82A and N88D/S occurring at frequencies of 0.5- 1% in four patients. Important to note was the presence of no more than one mutation per patient. UDP also detected minority RT mutations in each of the patients which were not detected by Sanger sequencing. Through this study they provided evidence for the frequency of PI DRMs at failure of PI based regimens. They

also showed the detection of minority RT DRM at PI regimen failure was indicative of persistence of RT mutations selected for at first-line regimen failure and their reversion to wild-type.

PR inhibitors have a high genetic barrier to resistance, and each primary PI DRM is associated with high reductions in viral PR fitness. These reasons could explain the observed infrequent detection of any (including minority) PI mutations at regimen failure in patients and why most patients fail to present with the signature PI DRM combinations. This is unlike nNRTIs that have a lower genetic barrier to resistance, and under drug pressure, DR variants are selected for and have been shown to result in virological failure.

1.12. Study rationale

By the end of 2018, there were over 5 million HIV-1 infected individuals on HAART in South Africa. The majority are on the first-line regimen, with approximately 200000 on second-line and 2000 on a third-line salvage regimen. Worldwide, population- based Sanger sequencing genotypic data has shown that the majority of patients failing a second line PI-based regimen do not harbour DRMs against PR inhibitors (155, 183-189). This was confirmed by a study conducted by our laboratory (152), which showed that only 58 of 350 (16.4%) individuals failing a second line PI-based regimen had DRMs in PR. Thus, there is a need to establish whether the presence of minority variants with PR DRMs contributed to treatment failure in this cohort. The availability of newer sequencing technologies allows for the detection of minority variants. Therefore, this study used NGS to detect the presence of DRM minority variants in the abovementioned South African cohort.

1.13. Aims and objectives

The overall aim of the study was to identify HIV-1 minority variants harbouring ARV DRM to PR inhibitors in a South African population failing a second line PI-based regimen and evaluate their impact on clinical outcome. This was achieved by the following objectives:

1) To RT-PCR amplify HIV-1 *pol* from 188 plasma samples from South African patients failing a second line PI-based regimen.
2) To detect drug resistant minority variants using NGS.

3) To evaluate the impact of the detected ARV drug resistant minority variants on clinical outcome.

Chapter 2

2.1. Samples used in this study

This study made use of stored participant samples collected from a previous project entitled 'Prevalence of antiretroviral drug resistance in patients failing Protease Inhibitor-based treatment: results from the first national survey in South Africa' (Steegen *et al.,* 2016). Study participants were recruited from eight South African provinces, including the Eastern Cape, Free State, Gauteng, Kwazulu-Natal, Limpopo, Mpumalanga, North West and Western Cape. Inclusion criteria included; minimum protease inhibitor-based ARV drug regimen duration of 6 months; participant had to be at least 18 years old; two successive viral loads above 1000 copies/ml (defined as virological failure) and a signed consent form.

For the purpose of this study, 188 stored plasma samples (stored at -80°C), with population-based Sanger sequences (*pol*), and associated antiretroviral drug resistance mutation profiles were available. Table 2.1 shows the study participant numbers available from each province; specimens from the Free State and Western Cape were not available due to storage at other laboratories. This study was conducted following approval from the University of the Witwatersrand Medical Human Research and Ethics Committee (HREC), with ethics clearance certificate number M1704127 (Appendix 1), sub- study certificate number M120254.

Table 2. 1: **The total number of participant samples from each of the six South African provinces available for this study.**

Province	**Sample number**
Eastern Cape (EC)	13
Gauteng (GP)	28
Kwazulu-Natal (KZ)	112
North-West (NW)	26
Mpumalanga (MP)	5
Limpopo (LP)	4
TOTAL	**188**

2.2. Next Generation Sequencing

HIV-1 *pol* from the 188 participant samples was sequenced using Next Generation Sequencing (NGS) on the Illumina Miseq® instrument, to characterise HIV-1 minority

variants. The NGS ARV drug resistance mutation profiles were compared to those available from the Sanger sequencing.

2.2.1. Viral RNA extraction

Total viral RNA was isolated from participant stored plasma, with a minimum input of 200µl (depending on how much was available) up to 500µl, using the EasyMag Extraction Kit (Biomerieux, INC, France) and NucliSeNS® EasyMag instrument (Biomerieux INC, France). Following an in- house protocol Sample was eluted in 25µl EasyMag elution buffer 3, as per manufacturer's instructions. Briefly, 500µl thawed participant plasma was added into each EasyMag cartridge well. Two mL of Lysis buffer was dispensed into each well as pre-determined by the automated system and incubated for 10 minutes. NucliSeNS EasyMag magnetic silica beads (50µl) were added and thoroughly mixed into each lysed sample. Following this, automated extraction continued as per manufacturer's instructions. RT-polymerase chain reaction (PCR) was performed immediately after RNA extraction.

2.2.2. RT-PCR Amplification

The partial *pol* gene, which consists of *PR* (coding for 6-99 amino acids) and *RT* (coding for 1-251 amino acids) was amplified using a two-step RT-PCR assay adapted from an accredited CDC genotyping assay(190). The RNA was first reverse transcribed into complementary DNA (cDNA) using Superscript™ III reverse transcriptase with platinum® Taq High fidelity (Invitrogen) and using outer forward primer PrtM-F1 (a 1:1 ratio of forward primers PrtM-F1a and PrtM-F1b) and outer reverse primer RT-R1 (see Table 2.2 for all primers used).

All master mix reagents used were prepared as detailed; 10µl Nuclease free water (Sigma, USA), 25µl 2X reaction mix, 2µl PrtM-F1 (10µM), 2µl RT-R1 (10µM) and 1µl Superscript III. Extracted viral RNA (10µl) was denatured at 65°C for 10 minutes and immediately cooled at 4°C for 1 minute. Reverse transcription was performed at 50°C for 45 minutes. The enzyme was the denatured at 94°C for 2 minutes. The following conditions were used to amplify the cDNA; denaturing of the DNA was performed at 94°C for 15 seconds; annealing of the primers at 50°C for 20 seconds and elongation at 72°C for 2 minutes, all repeated for 40 cycles. A final extension was completed at 72°C

for 10 minutes followed by a hold at 4°C. The second round of the amplification used AmpliTaq® Gold DNA polymerase kit (Thermofischer, Scientific, USA) with a DNA input of 4µl. The reagents combined for the mastermix as per single reaction were; 71µl nuclease free water (Sigma, USA), 10X AmpliTaq buffer at 10µl, 2µl dNTP (10µM), 8µl $MgCl_2$ (25µM), 2µl primer Prt-F2 (10µM); 2µl primer RT-R2 (10µM) and 1µl AmpliTaq Gold. The thermocycler conditions used were as follows; enzyme pre-activation was at 94°C for 4 minutes. 40 cycles of denaturation at 94°C for 15 seconds, primer annealing at 50°C for 20 seconds and primer elongation at 72°C for 2 minutes followed by a final extension step at 72°C for 10 minutes and a hold at 4°C.

Table 2.2: **Primer sequences used to amplify HIV-1 *pol* in an RT-PCR protocol. The nucleotide positions where each primer binds relative to the HXB2 reference sequence** (Accession number K03455) is also displayed**.**

Enzyme	Sequence	Nucleotide binding (bp)
PrtM-F1a	5'-TGA ARGAITGYACTGARAGRCAGG CTA AT-3'	1268-1296
PrtM-F1b	5'-ACT GAR AGR CAG GCT AAT TTT TTA G-3'	1279-1303
RT-R1	5'-ATC CCT GCA TAA ATC TGA CTT GC-3'	2559-2581
Prt-F2	5'-CTT TAR CTT CCC TCA RAT CAC TCT-3'	1454-1477
,RT-R2	5'-CTT CTG TAT GTC ATT GAC AGT CC-3'	2515-2537

2.2.3. PCR Product verification

The successful amplification of HIV-1 *PR/RT* genes was confirmed using agarose gel electrophoresis. Mass ruler DNA Ladder (80- 10 000bp) (Thermofischer Scientific, USA) was used as a molecular weight (MW) reference, in verifying that the correct amplicon with a band size of 1084bp was expressed.

Briefly a 2% agarose gel was prepared by dissolving 2g of agarose powder in 100ml of 2 % Tri- base buffer. GelRed® Nucleic Acid gel stain (Biotium, USA) was added to the agarose mixture at 10µl allowing for the DNA to be visualised under UV. This was then added to a gel tray and well impressions made using a gel comb. Once the gel had solidified, 10 µl mass ruler DNA ladder was loaded in the first well; then 5µl of each amplicon was mixed with 5µl 6X loading dye (Thermofischer Scientific, USA) and added into individual wells. The products were migrated for 45 mins at 100V and visualised on the Gel Doc system (Bio-Rad, USA).

2.2.4. DNA Purification

RT-PCR product was purified to remove any excess PCR components such as primers, dNTP's or buffer using the GeneJet Purification kit (Thermofischer Scientific, USA), which is a column-based purification system. The steps followed were as per manufacturer's instructions. Briefly, each amplicon volume was determined and added to an appropriately labelled 1.5ml microcentrifuge tube. GeneJet binding buffer was mixed into each tube in a 1:1 ratio. The mixture was transferred to a GeneJet purification column and centrifuged at 13000 X g for 1 minute. The resultant flow through was discarded. After this step the DNA was bound to the filter tube. An ethanol-based wash buffer (700µl) was added to the tube, which was centrifuged using the previously mentioned speed. The wash throughput was discarded, and the empty tube centrifuged once again to remove any remaining wash buffer. The purification membrane tube was removed and placed in a new 1.5ml microcentrifuge tube. Fifty µl of elution buffer was added to the filter membrane and subsequently centrifuged for 1 minute to unbind the purified DNA located at the bottom of the purification membrane. The purified DNA was collected in the microcentrifuge tube and stored at -20°C until further downstream processing.

2.2.5. DNA Amplicon Quantification

Purified HIV-1 DNA was quantified to determine the input DNA concentration prior to the library preparation and subsequent NGS, using the Qubit® dsDNA HS assay kit (Thermofischer Scientific, USA) on the Qubit®2.0 Fluorometer (Invitrogen). The Qubit® 2.0 uses fluorochromes that are specific for dsDNA and results in more accurate readings in comparison to the Nanodrop or UV-spectrometer which have less sensitivity and are affected by the presence of any other contaminants, including ssDNA and RNA. Briefly, a Qubit working solution was prepared to a 1:200 dilution by adding 1µl Qubit dye to 199µl Qubit buffer per sample measured. Subsequently, 198µland 190µl of the working solution were aliquoted into appropriately labelled Qubit tubes, to which 2µl of each DNA sample andten µl of each standard was added to the appropriate working solutionrespectively. The mixture tubes were vortexed for 3 seconds and incubatedfor 2 minutes at room temperature. DNA concentration measurements (ng/µl) were performed in triplicates and the average value recorded.

Thereafter, each DNA sample was appropriately diluted to a final concentration of 0.2ng/µl using molecular grade water (Thermo Fisher Scientific, USA). The calculation for the dilution was as follows:

[DNA concentration (ng/µl)] [Final volume stock solution to be added (2µl)]= Final Vol of solution

[Final concentration of new solution (0.2ng/µl)]

↓

Final volume of solution – DNA sample volume (2µl)

2.2.6. Nextera XT Library Preparation

The process of DNA library construction involves i) genomic DNA fragmentation, end repair and ligation of adapter oligomers (tagmentation); (ii) addition of universal sequences for sample indexing; (iii) cDNA fragment size selection and; (iv) normalization. These steps were all performed using the Nextera® XT DNA Library Kit (Illumina, USA) following the protocol as outlined in the Nextera library preparation manual (Illumina Nextera XT DNA Library preparation Reference manual, April 2017).

2.2.6.1. Tagmentation

The tagmentation process is outlined in Figure 2.1. Following dilution of viral cDNA (0.2ng/µl), 5µl of each sample (resulting in 1ng total concentration) was added to a 96-well microplate and followed by the addition of 10µl Tagmentation DNA (TD) buffer. Dual function Amplicon Tagment Mix (ATM), was added at a volume of 5µl to each well and the plate sealed using Microseal 'B' adhesive film. The reaction mixture was mixed using a rocking shaker, set at 400 X g, followed by centrifugation at a speed of 300 X g for 60 seconds. Tagmentation of the input DNA was completed by placing the reaction plate on the thermocycler for 5 minutes at 55°C, followed by a hold at 10°C. Immediately following completion of the tagmentation process, Tn5 transposomes still attached at ends of the fragment were removed and the reaction neutralised by addition of Neutralization Tagmentation buffer (NT Buffer). Five µl of NT buffer (at room temperature) was briefly vortexed and spun down; 5µl was added to each well. The reaction plate was covered with a new MicroSeal 'B' adhesive film and reaction contents

mixed as mentioned previously. The plate was kept at room temperature for 5 minutes to allow for the neutralisation reaction to reach completion. The plate was then placed on ice until the next PCR amplification step.

2.2.6.2 Adapter Index ligation

The tagmented adapter ligated DNA was amplified via a limited cycle PCR aimed at avoiding the development of chimera and heteroduplexes that bias fragment size estimation. This second round of PCR amplification adds index sequences that act as unique barcodes allowing for sequence multiplexing. Attached to each index are oligonucleotide sequences, P5 (forward adapter) and P7 (reverse adapter), which are complementary to those located on the flow cell of the Miseq, allowing for template strand to hybridize onto the flow cell.

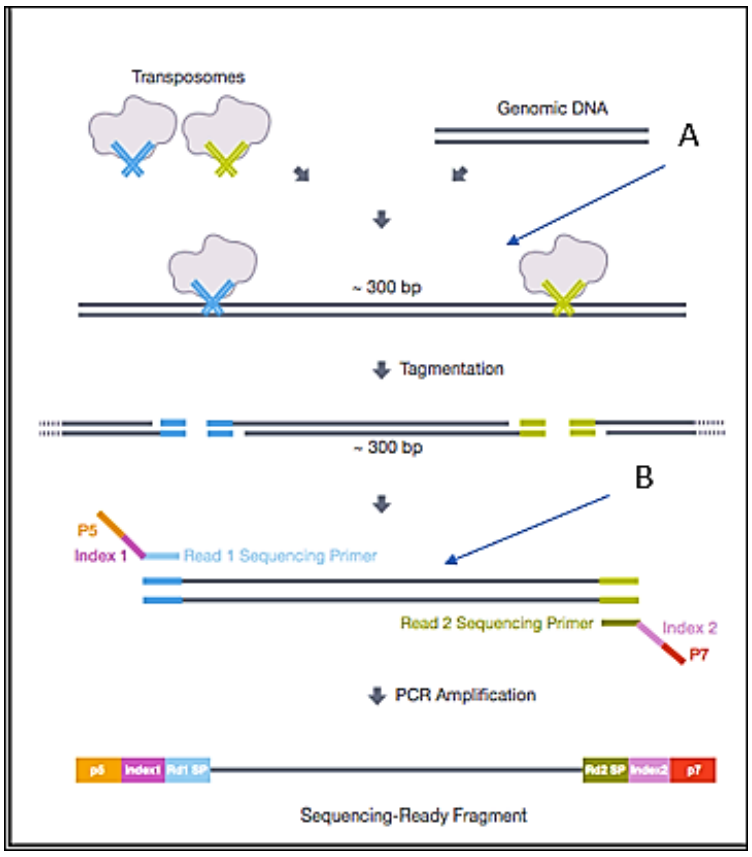

Figure 2.1: **The process of tagmentation using the Nextera XT DNA library preparation kit**. A) Input DNA is randomly fragmented by Transposomes resulting in fragmented DNA of average length of 300 bp while simultaneously ligating adapter sequences (blue and green) to either end of the fragment. B) Target sequence is barcoded using forward index 1 (dark purple) and reverse index 2 (light purple).

Attached to index 1 is a forward read adapter, P5 (orange) and attached to index 2 is reverse read sequence adapter P7 (red), which are used to hybridise the DNA to the flow cell for cluster generation and sequencing (copied from Illumina).

A sample plate sheet for the purpose of achieving correct index addition was generated using the Illumina Experimental Manager software (version 1.9.1). Thereafter, the reaction plate was removed from ice and 15µl NPM was added to each sample well of the reaction plate. This was followed by the addition of 5µl of each index primer (index primer 1 and 2) as indicated by the generated sample plate sheet. Following addition of these reagents, the final volume was 50µl. The plate was covered with a Microseal 'A' adhesive film and the reaction mixture was mixed using a rocking shaker at 400 X g for 1 minute and centrifugation at 300 X g for 1 minute. The following thermocycler conditions were used: 72°C for 3 mins, 95°C for 30 seconds followed by 15 cycles of: 95°C for 10 seconds, 55°C for 30 seconds, 72°C for 30 seconds, final extension at 72°C for 5 minutes and a hold at 10°C. The reaction plate was stored at -20°C until further downstream processing.

2.2.6.3. Fragment Purification and Size selection

The fragmented and barcoded samples were purified using the Agencourt® AMPure® XP kit (Beckman Coulter, USA), in a two-step bead precipitation process while simultaneously targeting fragments of 300bp - 600bp. The dual function AMPure beads are able to purify the DNA library by removing any adapter dimers and also allow for DNA fragment size selection.

First round AMPure® XP bead Precipitation

Before the purification process was commenced, solid phase reverse immobilisation (SPRI) AMPure® XP beads were allowed to stand at room temperature for approximately 30 minutes and all other reagents were thawed at room temperature. First round of the purification was performed using 45µl of the PCR mixture which was transferred appropriately to each well of a 0.8 ml storage plate. The AMPure XP beads were mixed well by vortexing for 30 seconds. A small amount of the AMPure XP bead solution was poured out in a Reagent Reservoir (VWR, USA). Twenty-five µl of the solution (calculation and theory for bead volume determination is shown in Figure 2.2) was carefully added and mixed into each well containing PCR mixture by pipetting up and down several times.

Selecting Fragment size

SPRI Agencourt® AMPure XP beads used for the purposes of size selection are based on a SPRI method in which magnetic beads bind DNA reversibly in the presence of polyethylene glycol (PEG) and NaCl (191). In a two-step precipitation, varying concentrations of PEG are used to achieve a narrow DNA size distribution. In the first round precipitation, a lower volume of bead solution/PEG concentration is used in order to remove all molecules larger than the upper limit of the desired size range (192). These molecules are dissolved in the supernatant and discarded. The beads are then resuspended in a resuspension buffer (RSB) which hydrates the DNA allowing it to be converted from an aggregated form back into solution. A higher concentration of bead solution is used in the second precipitation increasing the PEG concentration, thereby allowing for all DNA molecules with sequences longer than the lower limit of the desired interval to be removed (192). This all results in a narrow DNA fragment size distribution.

Bead size ratio

The target size distribution was 300–600bp. The Illumina Nextera tagmentation normally results in fragments of 200-500bp, this plus the added adaptors should give a size distribution of approximately 300- 600bp. In the first-round precipitation, a 0.55 X ratio determined using previous works of Connolly *et al*., 2010 was used (Figure 2.2). The volume of AMPure beads to be used was obtained by; ratio (0.55X) multiplied by volume of DNA solution (45 µl). Therefore, the volume of AMPure beads = 25µl.

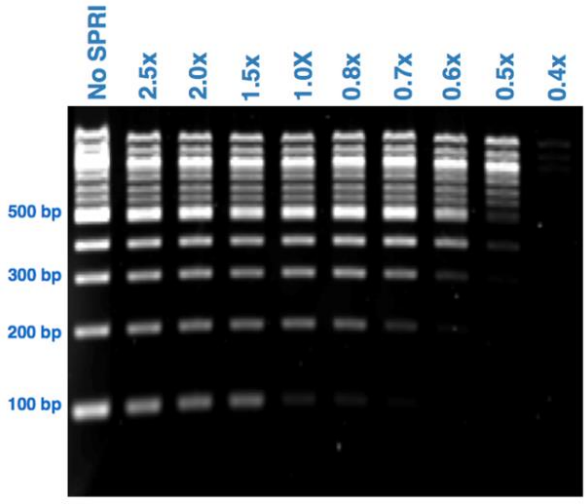

Figure 2. 2: **An agarose gel, obtained following fragmentation with different SPRI: DNA ratios. At higher SPRI: DNA ratios, large fragments were eliminated**. For example, at 0.7X ratio fragments of below 300bp were eliminated (copied from Connolly et al. (2010).

The reaction mixture was further mixed as previously mentioned and incubated at room temperature for 5 minutes. The plate was then placed on a 96 well magnetic stand (Ambion® RNA Life Technologies, USA) for a minimum of 2 minutes until the liquid colour had changed from brown to clear. The supernatant (63µl) was carefully removed with the plate still on the magnetic stand, making sure not to disturb the beads attached to the side of the well. With the plate still on the magnetic stand the beads were washed with 190µl 80% EtOH to remove any unbound reagents. The supernatant was removed and discarded. The EtOH wash step was repeated once again and supernatant discarded. Finally, the beads were allowed to air dry for 10-15 minutes to remove any residual EtOH. The reaction plate was removed from the stand and the beads with DNA library attached were each resuspended in 55µl of resuspension buffer (RSB) and the mixture was thoroughly mixed. The mixture was then incubated at room temperature for 2 minutes and then placed on the magnetic stand for another 2 mins or until the liquid turned to a clear colour. The resuspension (50µl) was carefully removed, paying special attention not to disturb the bead at the well wall and transferred to a new 0.8 ml storage plate.

Second round AMPure XP bead Precipitation

For the purposes of eliminating larger sized fragment molecules above ≥ 500bp, a 0.75X ratio was used. The volume of SPRI AMPure beads to be used in the second-round bead amplification was determined by multiplying 0.75X ratio by volume of solution (50µl), therefore, the volume of AMPure beads added was 37.5µl. The purified PCR mix was mixed with the beads by using the rocker shaker for 2 minutes at 400 X g. The mixture was further incubated for 5 minutes at room temperature. Following this, the plate was placed onto the 96 well magnetic stand until the liquid appeared clear in colour. With the plate still on the stand, 77µl of the supernatant was carefully removed and discarded. The beads, to which the DNA has been attached, were washed with 190µl of 80% EtOH and incubated for 30 seconds, and then 200µl of the supernatant was removed and discarded. The EtOH wash step was repeated and the beads allowed to air dry for 10-15 minutes, making sure any residual EtOH had evaporated. With the plate removed from the magnetic stand, 15µl of RSB was added to each well, the plate was sealed with a new MicroSeal ‘A’ film and reagents mixed on the rocking bioshaker (400 X g for 2 minutes). This was followed by incubation at room temperature for 2 minutes. The plate was then placed back on the magnetic stand until the liquid turned clear. Following completion of the purification, 10µl of the resuspension was transferred

to a new 96 well PCR plate, sealed with a new MicroSeal 'A 'film and appropriately labelled, day/month/year/purified library.

2.2.6.4. DNA Library Quantification

The concentration (ng/µl) of the purified library was determined as per manufacturer's instructions using the Qubit® dsDNA HS assay kit (Thermofischer Scientific, USA) on the Qubit®2.0 Fluorometer (Invitrogen) as previously detailed in section 2.2.5.

2.2.6.5. DNA Library Normalization and Pooling

After the DNA library was normalised to 2nM, each sample was pooled together in one microcentrifuge tube. This allows for multiplex sequencing to occur, as each DNA sample has a unique index identifier attached. Five µl of each normalised DNA library sample was transferred into a 1.5 mL centrifuge tube, and the pooled mixture was vortexed and centrifuged for 30s. The DNA library was denatured by mixing 5µl of the pooled DNA sample (2 nM) with 5µl freshly prepared (0.2 nM) NaOH, to give a final volume of 10µl. Briefly, the mixture was vortexed and centrifuged at 280 X g for 1 minute. The sample solution was incubated at room temperature for 5 minutes allowing for the denaturation to occur, and 990µl of pre-chilled HT1 was then added to the denatured DNA. This mixture was briefly vortexed and pulse- centrifuged resulting in a denatured DNA library concentration of 10pM. The DNA library was kept on ice before proceeding to the final dilution step.

Library Dilution

The 10pM denatured DNA library was diluted to a final volume of 600µl at a concentration of 7pM. In order to arrive at this final concentration, 10pM denatured DNA library at 420µl volume was added to a 1.5 microcentrifuge tube along with 120µl pre-chilled HT1 and 60µl of Phix control (10%). The DNA mixture was briefly vortexed and centrifuged then placed on ice until ready to load onto the MiSeq reagent cartridge.

2.2.6.6. Loading the DNA library onto the MiSeq reagent cartridge

The Illumina Miseq® Reagent Kit v2 and v3 were used on different sequencing runs. The version 3 kit which completes 600 cycles (300 cycles for forward and reverse

sequence) was used in the first run, whereas version 2 was used for the second run, since it provided sufficient coverage and was cheaper. The DNA library was loaded onto the cartridge and the MiSeq run was started in accordance with the Illumina MiSeq instrument on-screen instructions. Briefly, the flow cell was removed from the flow cell container, being careful to only touch the edges. The flow cell surface was rinsed with molecular grade water and dried by gently patting it with a lint- free lens cleaning tissue. This was followed by wiping with an alcohol wipe. All excess alcohol was dried with a lint free lens cleaning tissue. The flow cell was then loaded on the instrument as per the on-screen MiSeq control software (MCS) instructions. The DNA library sample was prepared for loading by using a fresh 1 ml pipette tip to pierce through the foil seal covering the 'loading sample' reservoir. Finally, 600µl of the DNA library sample was loaded into the reservoir and the cartridge was loaded onto the instrument when prompted by the on-screen instrument instructions.

Following the completion of a sequencing run, several QC criteria which included, cluster density (k/mm^2), passing filter and Q scores was analysed. The QC data was virtualised using the illumina Sequence Analysis viewer (SAV). Each metric is important for determining the quality of the run and overall reliability of the data.

Cluster density. The illumina Miseq V3 kit ideally results in a cluster density of 800 – 1200 K/mm^2. A run with a cluster density below 800 K/mm^2 is considered as poor and under clustered while cluster densities over 1200 K/mm^2 are considered over clustered. Over clustered flow cells lead to poor data resolution and interpretation with an inability to reliably distinguish between individual bases.

Clusters Passing Filter (PF): This is the number of clusters which passed the passing filter for each tile. A chastity filter removes the least reliable clusters from the image extraction with anything of a 0.6 score being filtered out. 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.

% Phasing: This is described as the average rate expressed as a percentage, at which individual molecules in a cluster become out of sync with each other. If molecules are falling behind and not being incorporated, it is described as pro-phres. Pre-phres occurs

when more than one nucleotide is incorporated per cycle and is referred to as jumping ahead.

Q score (Q30): This is the probability that a nucleotide has been incorrectly called. It is based on the number of reads correctly aligned to the phix control. Error rate is seen to increase towards the end of the sequence thus base sequence quality decreases towards the end of the sequence. A Q30 score represents a 1 in 1000 (0.1%) probability of a base being called incorrectly. Illumina instruments are associated with a Q30 of ≥ 70-80%.

2.3. Data Analysis

Following completion of the MiSeq runs, the resultant nucleotide sequence data was organised and analysed to identify HIV-1 variants using two bioinformatics pipelines, Geneious® software version 10.1 and the Deepchek® software. The NGS results from the two different programmes were compared to each other, as well as the available Sanger sequences.

2.3.1. Geneious

Sequencing data retrieved from the Illumina MiSeq run was analyzed using licensed Geneious version 10.1 software (https://www.geneious.com). The main steps involved in organizing the data and analyzing it for HIV drug resistance mutations are: 1) Pair forward (read 1) and reverse (read 2) pair end sequences for each sample; 2) align and trim the sequence ends of the sample sequences to an HIV reference sequence, HXB2 (GenBank accession # NC_K03455), and; 3) identify HIV drug resistance mutations in the RT and PR region based on the HBX2 HIV reference sequence. Briefly the HBX2 reference sequence was downloaded from Genbank. All known RT and PR drug class mutations were manually annotated onto the HBX2 HIV reference sequence to give an annotated reference sequence. Mutations were based on the published 2017 list of IAS-USA HIV drug resistance mutations (https://www.iasusa.org/resources/hiv-drug-resistance-mutations/).

Fastq files representing two paired-end reads per sample were imported into the Geneious program. The two sequence reads were paired to give one sequence per sample fastq file. These sequences were then aligned, and ends trimmed based on the

HBX2 reference sequence. These were then analyzed against the annotated reference sequence to identify HIV drug resistance variants. All nucleotide variants represented by at least 5 sequencing reads and at a frequency >1% from the reference sequence were then called using the variant finder. Variants and their frequencies were exported into an excel document and filtered for those present in amino acid sites known to correlate with ARV drug resistance as annotated on the reference sequence.

2.3.2. DeepChek®

Deepchek® software (ABL SA; https://www.ablsa.com/laboratory-solutions/deepchek-software) is a completely automated HIV data analysis pipeline. The licensed software was accessed at the National Institute for Communicable Diseases, Johannesburg, South Africa. Two raw pair end fastq files (forward and reverse sequences) per samples were uploaded onto the Deepchek® software, with a maximum of 10 samples being uploaded and processed at any one time. The software automatically paired the sequence reads, aligned and trimmed the ends using an HIV reference sequence, filtered the sequences to eliminate low quality reads (system coded) and identified HIV sequence variants based on the Stanford HIV drug resistance database (https://hivdb.stanford.edu). A clinical report giving a summary of all variants identified at a minimum threshold of 1% was produced and saved on the online database storage cloud.

2.3.3. Phylogenetic Analysis

To ensure that the Sanger and NGS sequences were the same for each participant, as well as for subtyping, a multiple alignment of the *pol* region with references from HIV-1 subtypes A to K, CRF01_AE and CRF02_AG (http:=hiv-web.lanl.gov) was generated in Clustal X2 (193)). Aligned sequences were converted to MEGA Version 10.0.4 format and used in phylogenetic and molecular evolutionary analyses(194). Phylogeny reconstruction of each *pol* gene was performed by neighbour-joining using the Kimura two-parameter distance matrix, and stability of the nodes was assessed by bootstrap analysis (1000 replicates). Bootstrap values >70% were considered significant.

2.4. Statistical analysis

To access the level of significance in differences of mutation numbers detected between Sanger sequencing and NGS we conducted p- test using GraphPad (v.8.0)(195) available online https://www.graphpad.com/quickcalcs/ttest1.cfm. All statistical analysis relating to participant demographics was performed using Microsoft Excel. The level of mutational correlation (mutation frequency ≥ 20%) between NGS and Sanger was determined.

Chapter 3

3.1. Participant demographic and clinical data

The demographic and clinical data of the 188 participants included in this study are shown in Table 3.1. The majority of the participants (72.9%, n= 137) were female. The median VL was 4.8 log_{10} (IQR: 4.2-53) RNA copies/ml, and most participants (56.2%, n= 104) were on a TDF- 3TC-LPV/r regimen. Participants were on their protease inhibitor-based regimens for a median time of 23 months (IQR, 12.5-35.0 months). Based on the available Sanger sequences, all participants were infected with HIV-1 subtype C. Appendix 2 contains the detailed demographic and clinical characteristics for all participants.

Table 3.1: **Demographic and clinical characteristics for the 188 HIV-1 positive participants failing protease inhibitor based antiretroviral drug regimens included in this study from all six provinces across South Africa.**

Characteristics	**Number of participants**	**Value**
Gender (Female)	137	
Age (mean; SD) (**years**)	188	38.4; 9.9
CD4 cell count regimen (median; min-max; IQR) (**cells/mm3**)	175	190.0; 5.0-984.0; 91-287.3
HIV VL regimen (median; min-max; IQR) (**log10 RNA copies/ml**)	184	4.8; 3.0-6.5; 4.2-5.3
Time on PI regimen (median; min-max; IQR) (**months**)	173	23.0; 6.0-105; 12.5-35
Regimens		
TDF-3TC, LPV/r	104	
AZT-3TC-LPV/r	54	
AZT-ddI-LPV/r	10	
ABC-3TC-LPV/r	7	
D4T-3TC-LPV/r	4	
AZT-TDF-3TC-LPV/r	4	
Other	5	

3.2. RT-PCR amplification of *PR/RT*

Following viral RNA extraction, the RT-PCR amplification was successful for 158 of the 188 (84.0%) participant samples, as confirmed by agarose gel electrophoresis (Figure 3.1). The expected size of approximately 1084bp was amplified, and amplicons were

successfully purified and subsequently used in the NGS protocol. Figure 3.2 highlights (in red) the geographic distribution of the participants from which the 158 samples were amplified, with the majority coming from Kwazulu-Natal.

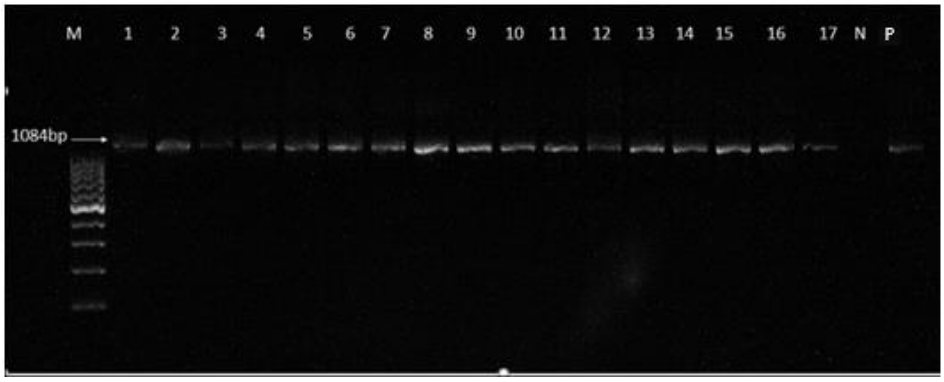

Figure 3. 1: **Agarose gel (2%) electrophoresis showing successful amplification of HIV-1 *PR/RT* amplicons at the expected size of 1084bp (indicated by arrow**).
M= 100-1000bp Molecular weight DNA ladder (Thermofischer Scientific); N=negative control; P= positive control; Lanes 1-17= amplicons from 17 participants.

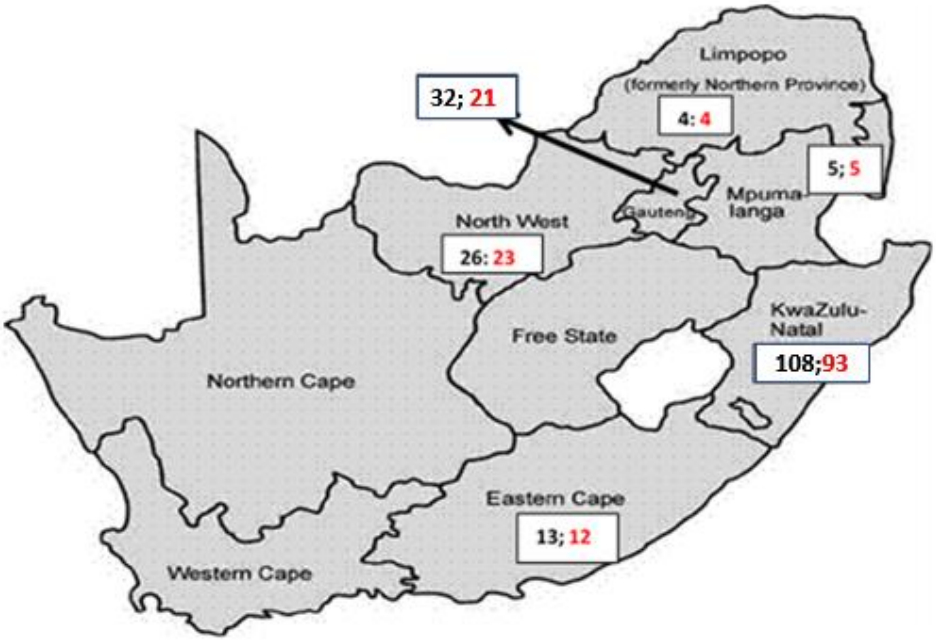

Figure 3. 2: **South African map showing the numbers of HIV-1 positive participant samples available (n=188; black text) and the number that was successfully RT-PCR amplified (n=158; red text) in the relevant provinces across South Africa**.

3.3. Illumina MiSeq NGS

3.3.1. NGS quality metrics

The 158 purified amplicons were successfully sequenced on the Illumina MiSeq. The sequencing metrics for the two MiSeq run results including the read quality of each run are shown in Table.3.2. Overall, the % error rates were low, and the quality scores indicated that all the reads were of sufficient quality to obtain usable data, with run two yielding better scores. The percentage of reads from clusters in each tile that are aligned to the PhiX genome are represented by the % aligned. This provides an indication on the success of the clustering process, e.g. higher values may signify underclustering or imprecise cluster density prediction.

Table 3. 2: Miseq sequencing metrics recorded for the two sequencing runs.

Level	Yield Total (G)	Projected Total Yield (G)	Aligned (%)	Error Rate (%)	Intensity Cycle 1	% >= Q30
Run 1						
Read 1	**6.34**	**6.34**	**18.79**	**2.68**	**126**	**65.80**
Read 2	**0.15**	**0.15**	**0.00**	**0.00**	**255**	**82.31**
Read 3	**0.15**	**0.15**	**0.00**	**0.00**	**190**	**96.28**
Read 4	**6.34**	**6.34**	**18.40**	**4.43**	**118**	**57.95**
Non-Indexed Total	**12.67**	**12.67**	**18.60**	**3.56**	**122**	**61.87**
Total	**12.97**	**12.97**	**18.60**	**3.56**	**172**	**62.50**
Run 2						
Read 1	**4.46**	**4.46**	**44.43**	**1.29**	**105**	**88.50**
Read 2	**0.16**	**0.16**	**0.00**	**0.00**	**110**	**61.56**
Read 3	**0.16**	**0.16**	**0.00**	**0.00**	**177**	**96.38**
Read 4	**4.46**	**4.46**	**42.44**	**1.81**	**98**	**80.73**
Non-Indexed Total	**8.93**	**8.93**	**43.44**	**1.55**	**101**	**84.62**
Total	**9.24**	**9.24**	**43.44**	**1.55**	**122**	**84.43**

% >= 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

3.3.2. NGS Sequence analysis

The NGS data obtained from both Miseq runs was then extensively analysed using the Geneious® and Deepchek® software, and any participant samples containing variants associated with known PI and RT mutations were evaluated and software results compared to each other as well as to the known Sanger sequences (Appendix 2).

Analysis of mutations detected using Geneious® or Deepchek® showed comparable results for variants that were present at proportions ≥ 4.5%. However, differences were noted at levels less than 4.5%, therefore, for all subsequent analyses, a cut off of 4.5% was used. Mutations present at proportions of 4.5 to 20% were considered minority variants.

3.3.3. Antiretroviral drug resistance mutations

3.3.3.1. Protease Inhibitor mutations

Variant sequence analysis from Geneious® and Deepchek® confirmed that 26 of the 158 samples contained at least one major or accessory mutation associated with PI resistance. The available Sanger sequences confirmed that 24 (15.2%) participants harboured HIV-1 with major or accessory PI DR mutations (Table 3.3), all of which were detected by NGS. Thus, NGS detected an additional two participants (NWC005 and GPC032) that harboured HIV-1 with PI DR mutations. Mutations in these two participants were present as minority variants: in participant NWC005, the accessory mutation K20T was detected at 4.7%; whereas in GPC032, the major mutations M46I, I54V, V82A and L90M were all detected at below 6.4%.

Looking at the NGS sequence data, seven of the 26 participants harboured HIV-1 with at least one major PI resistance mutation, 13 had a combination of major and accessory PI mutations, and six participants presented with accessory PI resistance mutations only (Table 3.3).A comparison of the drug resistance levels relevant to the PIs available in the South African public sector, as determined by Sanger versus NGS data, is also highlighted in Table 3.3. For three of the samples minority variants were detected in addition to those also detected by Sanger sequencing. Only two of the five samples which contained minority PI variants, as detected by NGS, had an impact on the drug resistance scores. Inclusion of these minority variants in the analysis resulted in a

change from susceptible to high-level resistance for at least one drug for sample GPC032, and a change from potential low-level resistance to high-level resistance levels for sample MPC006, For participant MPC006 the detection of major variants V32I, I47V, I54V, V82A, L76V and accessory K20T otherwise undetected by Sanger sequencing increased the level of resistance to all three PI's from potential low-level resistance to high-level resistance (Table 3.3). In participant GPC032, detection of M46I, I54V and V82A, all unrecognised by Sanger sequencing, yielded high-level resistance levels to ATV/r and LPV/r. By contrast, for the other three participants (KZC140; NWC005; LPC005) there was no change in the genotypic resistance level

NGS identified all mutations detected by Sanger sequencing, with the exception of major mutations I54V and I47V detected by Sanger in samples KZC121 and NWC016, respectively. This resulted in an under scoring of the genotypic resistance to PI drug ATV//r in the former while no change in the resistance was seen in the latter (Table.3.3).

Table 3. 3: PI HIVDR mutations and their cumulative genotypic resistance scores as detected by Sanger sequencing and NGS in 26 participants within the cohort. Shown in bold are major PI DR mutations detected in each participant, along with five participants with PI minority variants detected (asterisk). Sanger detected mutations not recognised by NGS in two participants as shown with orange highlight.

			ARV DRUGS		
Participant ID	**Sequencing Method**	**Deepchek® (%); Geneious® (%)**	**LPV/r**	**ATV/r**	**DRV/r**
KZC121	Sanger	L10F, **I54V, L76V, V82A**	HLR	IR	LLR
	NGS	L10F (96.4; 96.5), (L76V (98.9;99.9), **V82A** (99.3;98.5)	HLR	LLR	LLR
NWC016*	Sanger	**M46I, I47V, I54V, V82A**	HLR	HLR	S
	NGS	**M46I** (93.7;93.7), **M46L** (6.1;6.1), **I54V** (99.9;99.1), **V82A** (94.6;94.6), **V82S** (5.3;5.3)	HLR	HLR	S
GPC032*	Sanger	ND*	S	S	S
	NGS	**M46I** (4.5;4.4), **I54V** (3.7;4.7), **V82A** (4.6;6.3), **L90M** (5.5;5.6)	HLR	HLR	S
KZC140*	Sanger	**I54V**	LLR	LLR	S
	NGS	**I54V** (20.7;21.4), **M46I** (16.2;20.1)	LLR	LLR	S
	NGS	L24I (50.1;0.0), L33F (56.7;0.0) **M46I** (54.0;64.8), **I54V** (55.7;64.6), **V82A** (61.4;67.7), L10F (64.0;66.2)	HLR	HLR	PLLR
MPC006*	Sanger	**M46I**	PLLR	PLLR	PLLR
	NGS	**M46I** (22.7;18.7), K20T (18.3;18.3), **V32I** (6.2;0.0), **I47V** (18.1;18.3), **I54V** (4.5;3.3), **V82A** (6.1;5.7), **L76V**	HLR	HLR	HLR
NWC005*	Sanger	ND*	S	S	S
	NGS	K20T (4.7;4.6)	S	S	S
ECC018	Sanger	**I54V,** Q58QE**, L76V**, **V82A**	HLR	LLR	LLR
	NGS	**I54V** (83.0;85.7), Q58E (55.0;60.0), **L76V** (91.0;97.3) **V82A** (67.0;66.6)	HLR	LLR	LLR
GPC014	Sanger	**L90M**	LLR	LLR	S
	NGS	**L90M** (94.0;98.3)	LLR	LLR	S
KZC012	Sanger	K20T	S	S	S

	NGS	K20T (99.0;100.0)	S	S	S
KZC034	Sanger	K20T, L23LI**, M46I, I47V, L76V**	HLR	LLR	IR
	NGS	K20T (99.6;100.0), L23I (46.9) **M46**I (99.7;100.0), **I47V** (99.3;0.0), **L76V** (99.3;99.9)	HLR	LLR	IR
KZC060	Sanger	**M46ML, I54V, V82VA**	HLR	HLR	S
	NGS	**I54V** (99.41;99.9), **V82A** (37.3;40.1) **M46I (**54.4;58.0), **M46L** (25.37;21.0)	HLR	HLR	S
KZC076	Sanger	**L10F I54V,L76V, V82A**	HLR	IR	LLR
	NGS	**I54V** (98.0;99.9), **V82A** (96.0;94.6), L10F (97.0;97.0), **L76V** (97.0;97.0)	HLR	IR	LLR
KZC092	Sanger	L10F, K20T, L24I, **M46I, I54V, V82A,L76V**	HLR	HLR	LLR
	NGS	**L10F (99.0;100.0), K20T** (76.0;80.1), L24I (24.0;31.2), **M46I** (99.0;99.6), **I54V** (99.0;99.6), **V82A** (98.0;98.0), **L76V** (98.0;98.1)	HLR	HLR	LLR
KZC101	Sanger	M46L, I54V, V82A	HLR	HLR	S
	NGS	M46L (99.3;99.9), I54V (99.0;100.0), V82A (99.3;99.9)	HLR	HLR	S
KZC126	Sanger	L10F, L33F, **M46L, I50V, I54V, V82A**	HLR	HLR	IR
	NGS	L33F (98.;99.9), **M46L** (98.0;99.9), **I54V** (99.0;99.9), L10F (98.0;98.3), **I50V** (96.0;99.9)**, V82A** (97.0;99.7)	HLR	HLR	IR
KZC135	Sanger	K20KT, **V32VI**, **M46I**, L10F, **I54V**, **V82A**	HLR	HLR	LLR
	NGS	K20T (35.9;38.6), **V32I** (49.2;0.0), **M46I** (99.2;99.4), **I47V** (53.7;58.3**),** **V82A** (36.4;37.4), L10F (33.7;38.0),	HLR	HLR	LLR
KZC147	Sanger	**M46I, I54V, L76V, V82A**, L10F	HLR	HLR	LLR
	NGS	**M46I** (98.1;98.5), **I54V** (99.3;99.8), **L76V** (99.0;98.5), **V82A** (99.4;99.5) L10F (99.1;100.0)	HLR	HLR	LLR
KZC156 ,	Sanger	Q58E	S	S	S
	NGS	Q58E (92.0;92.7)	S	S	S

KZC157	Sanger	L33F	S	S	S
	NGS	L33F (99.3;100.0)	S	S	S
KZC158	Sanger	**L10F, V82A**,	IR	LLR	S
	NGS	**V82A** (96.4;96.9), L10F (81.9;83.0)	IR	LLR	S
KZC183	Sanger	Q58E	S	S	S
	NGS	Q58E (61.4;59.3)	S	S	S
LPC005	Sanger	**M46I, I54V, V82A**, L10F, L24I, L33F	HLR	HLR	PLLR
	NGS	L24I (82.7;0.0) L33F (82.4;0.0) **M46I** (99.5;99.9), L10F (88.2;88.4) **I54V** (99.4;99.9), **V82A** (99.4;99.8)	HLR	HLR	PLLR
LPC006	Sanger	**I54V, V82A**	IR	IR	S
	NGS	**I54V** (99.2;99.8), **V82A** (99.2;99.6)	IR	IR	S
NWC003	Sanger	L10LF	S	S	S
	NGS	L10F (94.0;96.9)	S	S	S

S Susceptible;PLLR Potential Low-Level Resistance ;LLR Low-Level Resistance; IR Intermediate Resistance HLR High-Level Resistance;*ND: No mutations detected; 0.0%= No detection by software

Overall, the most frequent major PI resistance mutations detected amongst the 26 participant samples were V82A, followed by I54V and M46IL. It is worth noting that, these mutations were also the most abundant among minority variants (Figure 3.3). The same numbers of participants were recorded from the Geneious® analysis software, however the proportions vary slightly (results not shown). Interestingly, 12 participant samples had a mutation combination of M46IL, I54V and V82A (Table 3.3 and Figure 3.3). The most common accessory PI mutation noted was L10F.

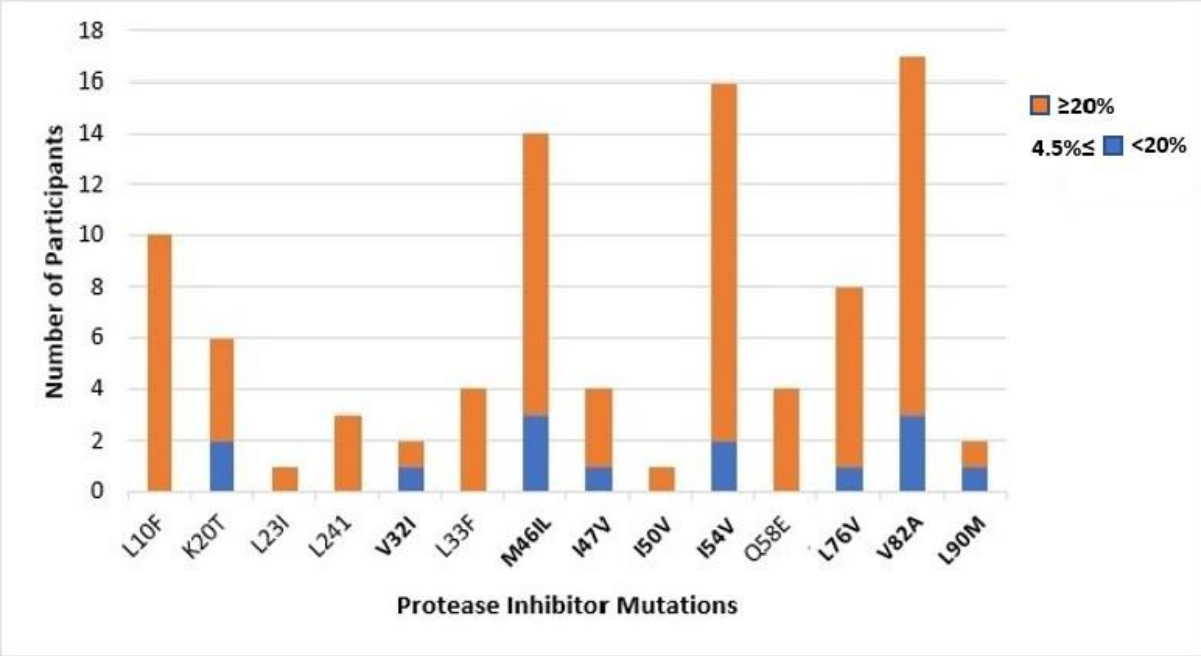

Figure 3. 3: **Frequency of major and accessory Protease inhibitor (PI) drug resistance mutations as detected by NGS and analysed by Deepchek® for 26/158 participants**. The major PI mutations are highlighted in bold.

3.3.3.1.1. Phylogenetic analysis

Phylogenetic analysis was performed in order to ensure genetic relatedness of the NGS and Sanger sequences for each of the 26 samples harbouring PI mutations, and additionally to confirm the subtype. A consensus sequence for each sample from the NGS data was obtained and used in phylogenetic tree analysis, together with the available population-based Sanger sequences (Figure 3.4). All 26 samples were confirmed to be HIV-1 subtype C and each matched Sanger and NGS sequence clustered together, excluding cross-contamination.

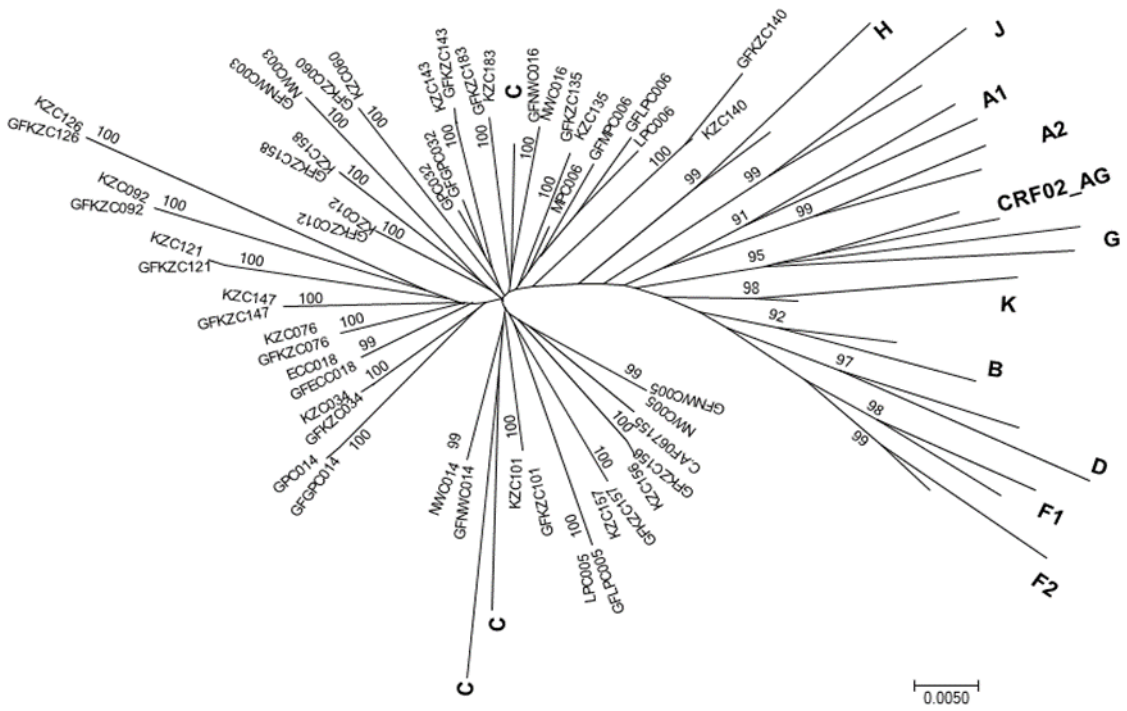

Figure 3. 4: Neighbour joining tree of matched NGS and Sanger sequences from each of the 26 participants harbouring HIV-1 with protease inhibitor drug resistance mutations.

3.3.4. RT mutations

The NGS data was further analysed for the presence of DR mutations to five currently available RT inhibitors in the South African ART program, ABC, AZT, TDF, 3TC/FTC and EFV**.**

Overall, HIV RT inhibitor resistance was detected in 130 of the 158 sample sequences by NGS. Of these, single drug class NRTI and NNRTI DR mutations were detected in 16 (10.1%) and 10 (6.3%) of the participants, respectively, while 104 harboured a combination of mutations impacting both classes (Appendix 3). The available Sanger data confirmed RT inhibitor mutations in 123 of the participant samples, all of which were detected by NGS. Therefore, NGS detected mutations in seven additional samples (Appendix 3).

Interestingly, 46 of the 130 sample sequences had additional minority HIV-1 RT inhibitor variants detected and confirmed by both Geneious® and Deepchek®. These resulted in changes in the reported resistance levels for 29 of the participant samples (Table 3.4). The mutations for the remaining 17 are shown in Appendix 3. Three of the 29 participants had resistance to at least one additional RT inhibitor and for 26 participants, higher genotypic resistance levels was reported for at least one RT inhibitor (Table 3.4). For example, participant KZC018 harboured minority variants, M184VI at proportions less than 18%, which were otherwise undetected by Sanger sequencing. This resulted in low-level resistance to ABC and high-level resistance to 3TC/FTC. Additionally, NGS detected variant D67G at proportions of 10.3% in participant KZC130, which led to a change in resistance to AZT, from potential low-level resistance to high-levels of resistance.

Moreover, RT inhibitor minority variants were detected in eight of the samples with PI DR mutations. For three of the eight participant samples there was a change in the levels of genotypic resistance to at least one RT inhibitor drug. Sample MPC006 (Table 3.4), had minority mutants D67N, K70R, A98G, E138K and K238T which resulted in high-level resistance to ABC and low-level resistance to both AZT and TDF. Again, detection of K70E, M184I and T215I in EC018, led to low-level resistance to AZT and TDF and intermediate resistance to AZT. In the third participant KZC140, additional

detection of K70R, Y181C, G190A and H221Y resulted in low-level resistance and high-level resistance to AZT and EFV, respectively (Table 3.4).

Overall, the most prevalent RT inhibitor DR mutations were NRTI associated M184VI, followed by NNRTI associated K103NS, E138AK and D67G (Figure 3.4). In the case of minority RT inhibitor DR variants NNRTI associated variants K103NS, D67E and E138AG were detected most frequently (Figure 3.4).

Table 3. 4: Genotypic resistance profiles for the 29 participants with HIV-1 harbouring minority RT inhibitor mutations which results in a change in resistance levels to at least one ARV drug. Participant samples with minority HIV variants affecting resistance levels for all three drug classes are shown in asterisk**.**

Participant ID	Sequencing Method	RT mutations		ARV drugs				
		Deepchek® (%); Geneious® (%)		ABC	TDF	FTC	AZT	EFV
		NRTI	NNRTI					
ECC006	Sanger	M184V	K101HKQ, V106M, G190A,	LLR	S	HLR	S	S
	NGS	M184V (98.0;99.9),	K101E (10.0;5.8), K101H (86.0;94.1), V106M (97.0;99.9), G190A (88.0;92.5)	LLR	S	HLR	S	HLR
ECC013	Sanger	M41ML, A62AV, K65R, M184V,	K219HQN	HLR	HLR	HLR	S	S
	NGS	M41L (27.0;0.0), A62V (32.0;0.0), K65R (26.0;0.0), D67N (15.0;0.0) M184V (98.0;0.0)	ND*	HLR	HLR	HLR	LLR	S
ECC018*	Sanger	D67N, M184V	ND*	S	S	HLR	S	S
	NGS	D67N (28.0;32.6), K70E (11.0; 12.3) M184V (54.0;55.0), M184I (15.0; 16.4) T215I (5.0;3.9)	ND*	LLR	IR	HLR	IR	S
GPC015	Sanger	ND*	K103N	S	S	S	S	HLR
	NGS	D67E (8.0;0.0) F77L (5.0;0.0),	K103N (86;8;85.1), V106M (6.0;6.2)	PLLR	PLLR	S	LLR	HLR
GPC058	Sanger	E138A	ND*	S	S	S	S	S

	NGS	E138A (97.0;99.4); K65R (11.0;10.0)	ND*	IR	HLR	IR	S	S
KZC005	Sanger	K65R, M184V,	K101E, V106M, E138A, G190A,	HLR	IR	HLR	S	HLR
	NGS	K65R (58.0;62.0), M184V (85.0;91.0), D67N (14.0;14.9), L74V (18;17.2)	K101E (75.0;81.8), V106M (76.0;86.1),G190A (92.0;95.2) E138A (89.0;93.0), M230L (4.0;4.6), K238N (5.0;0.0)	HLR	HLR	HLR	LLR	HLR
KZC018	Sanger	ND*	ND*	S	S	S	S	S
	NGS	M184I (6.4;3.0), M184V (17.1;11.8)	ND*	LLR	S	HLR	S	S
KZC023	Sanger	D67N	ND*	S	S	S	S	S
	NGS	D67N (20.0;0.0)	K103N (9.0;7.2), V106M (14; 7; 16.8.)	S	S	S	LLR	HLR
KZC040	Sanger	ND*	K101H, V108I, Y181C, G190A	S	S	S	S	HLR
	NGS	M184V (21.0;10.8),	Y181C (25.0;11.1) G190A (25.0;13.9),	S	S	HLR	S	HLR
KZC044	Sanger	ND*	ND*	S	IR	S	S	S
	NGS	K65R (11.0;13.4),	E138A (30;25.2)	IR	HLR	IR	S	S
KZC054	Sanger	ND*	ND*	S	S	S	S	S
	NGS	ND*	V106(9.0;5.8), V179D	S	S	S	S	HLR
KZC055	Sanger	K70R, V75M, F77L, M184V	K101H, K103N, G190A	LLR	S	HLR	IR	S
	NGS	M41L (23.9;25.1), K65R (12.6;13.2), K70R (68.20;74.4), F77L (99.3;100), V75M (99.3;99.9) M184V (99.3;99.9)	K101H (99.2;99.6), K103N (98.4;99.8), G190A (99.5;99.9)	HLR	HLR	HLR	IR	HLR

KZC059	Sanger	ND*	L100I, K103N	S	S	S	S	S
	NGS	ND*	L100I (98.5;99.3), K103N (98.6;99.4) T215A (9.5;7.8)	S	S	S	LLR	HLR
KZC084	Sanger	A98G	ND*	S	S	HLR	S	S
	NGS	D67N (12.8;13.6) A98G (73.2;78.6)	K103N (17.1;15.0), Y318I (23.5;22.0)	LLR	S	HLR	S	S
KZC093	Sanger	K70R, A98G M184V	K103N, P225H, T215FITIS	LLR	S	HLR	LLR	HLR
	NGS	K70R (20.8;18.7), M184V (99.0;94.2), A98G (94.3;95.8).	K103N (80.1;80.0), K103S (19.4;19.8), T215F (9.4;12.3), K219Q (8.0;6.3) P225H (86.8;87.0)	IR	LLR	HLR	HLR	HLR
KZC094	Sanger	M184V	K103N, Y181C	S	S	HLR	S	HLR
	NGS	D67N (16.0;16.9), M184V (15.0;0.0)	K103N (54.0;56.8), K103S (11.0;11.7), V108I (5.0;7.4), Y181C (74.0;75.6), H221Y(13.0;14.6)	LLR	S	HLR	S	HLR
KZC112	Sanger	ND*	ND*	S	S	S	S	S
	NGS	M184V (20.97;16.3)	K103N (15.16;13.9)	LLR	S	HLR	S	HLR
KZC130	Sanger	V75M, F77L, M184V	K103N, Y188C	LLR	S	HLR	PLLR	HLR
	NGS	A62V (89.7;89.7), D67G (10.3;9.3), V75M (98.9;99.8), F77L (98.9;99.9), M184V (98.8;99.9)	K103N (98.3;100), Y188C (98.6;99.9)	LLR	S	HLR	LLR	HLR
KZC133	Sanger	M184IV	ND*	S	S	HLR	S	S

	NGS	K65R (9.9;8.4), M184I (30.6;23.3), M184V (43.8;52.5)	K103N (10.0;8.6)	LLR	S	HLR	S	HLR
KZC139	Sanger		V106M, V108I	S	S	S	S	S
	NGS	M41L (5.0 ; 5.4), M184V (54.0 ; 57.2)	V106M (87.0 ; 94.7), V108I (74.0 ; 83.3) Y188F (10.0 ; 7.5)	LLR	S	HLR	S	HLR
KZC140*	Sanger	M184V	ND*	LLR	S	HLR	S	S
	NGS	K70R (13.0;11.9) M184V (55.6; 56.1),	Y181C (25.5; 24.3), Y188F (10.0 ; 0.0),) G190A (20.8;20.0), H221Y (16.2; 16.3)	LLR	S	HLR	LLR	HLR
KZC143	Sanger	M184V	K103S	LLR	S	HLR	S	HLR
	NGS	M184V (99.6;99.8), T215Y (10.5;0.0)	K103S (21.6;21.0), G190A (9.4;0.0)	LLR	S	HLR	IR	HLR
KZC148	Sanger	M184V, A98G	V179D	LLR	S	HLR	S	S
	NGS	F77L (20.0) A98G (99.0;100.0), M184V (98.0;100)	ND*	LLR	S	HLR	S	HLR
KZC156	Sanger	M184V	K103N, Y318F	S	S	HLR	S	S
	NGS	M184V (92.0;93.4)	K103N (90.0;99.0), Y318F (100.0;.0), E138Q (19.0;17.8)	S	S	HLR	LLR	HLR
KZC183	Sanger	ND*	V106M, G190GA	S	S	S	S	S
KZC193	Sanger	D67N, M184V	K101HG	S	S	S	S	S
	NGS	D67N (85.6;88.1), A98G (38.6;43.1), K101E (22.2;18.2) M184V (87.1;99.5)	K101H (35.1;32.7) G190S (13.6;11.1)	LLR	S	S	S	HLR
KZC195	Sanger	ND*	ND*	S	S	S	S	S

	NGS	M41L (7.1;7.4), M184V (12.0;10.3)	V179E (9.3;8.0)	LLR	S	HLR	S	PLLR
MPC006*	Sanger	M184V	K103N.P225H	LLR	S	HLR	S	HLR
	NGS	D67N (15.5;14.0), K70R (14.7;13.2) A98G (10.9;8.8) M184V (53.9;53.1), K219E (13.8;17.9)	K103N (73.6;77.1), E138K (17.2;15.4), P225H (44.97;41.0) K238T (14.6;15.0)	HLR	LLR	HLR	IR	HLR
NWC019	Sanger	M184V	Y188L	LLR	S	HLR	S	HLR
	NGS	M184V (98.0;100.0), D67E (5.0; 4.1)	Y188L (96.0;97.3)	LLR	S	HLR	LLR	HLR

S Susceptible; PLLR Potential Low-Level Resistance; LLR Low-Level Resistance; IR Intermediate Resistance; HLR High-Level Resistance;*ND: No mutations detected; 0.0%= No detection by software

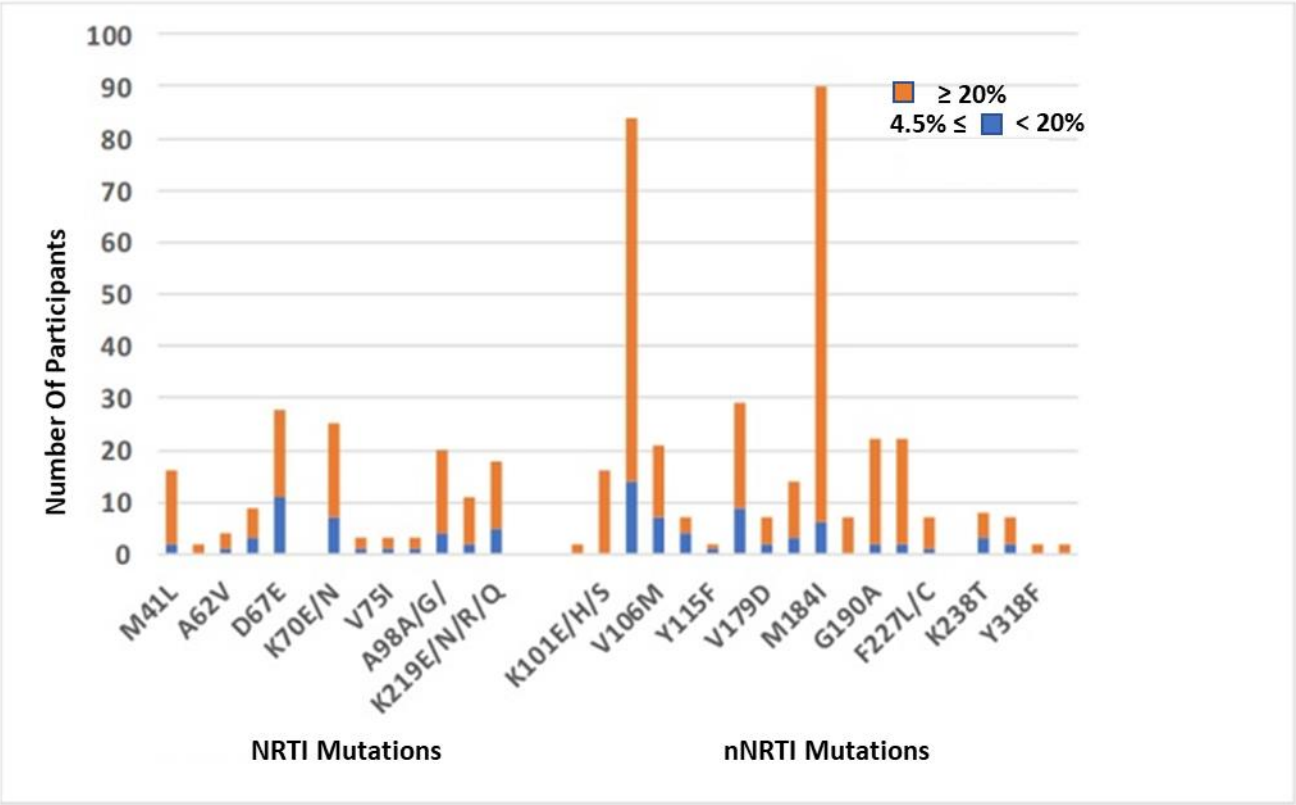

Figure 3.5: Frequency of reverse transcriptase inhibitor drug resistance mutations detected at proportions ≥4.5 and at ≥ 20 in 130 of the 158 participants included in the study.

Chapter 4

4. Discussion

ARV drug resistance genotypic testing using population based Sanger sequencing for HIV-1 infected patients that are virologically failing a second-line PI based drug regimen has infrequently detected HIV-1 DRMs in PR (152, 153, 196, 197). It has been suggested that the presence of PI DR minority variants that are generally undetected by population-based Sanger sequencing could possibly contribute to an increased resistance to PIs and subsequent virological failure.

Thus, this study aimed to determine whether the presence of previously undetected minority variants encoding for PI drug resistance contributed to the virological failure of HIV-1 infected South Africans on a second-line PI based regimen.

4.1. HIV-1 RNA extraction and amplification

Of the 350 participant samples from HIV-1 infected South Africans failing a second-line PI based regimen (152), a total of 188 samples were available for the purposes of this study. For 19.0% (n=30) of the 188 participants for which multiple RT-PCR amplification attempts failed, we hypothesise that multiple freeze-thaw cycles could have resulted in a decrease in the level of RNA integrity. It has been shown that even ≥ 3 freeze/thaw cycles are enough to decrease the level of RNA integrity by 35% (198, 199). Additionally, some of the samples were grossly haemolysed (visibility of a dark red colour), a condition which implies inappropriate transport of samples and/or isolation and storage of plasma. Panaccio *et a*l., (1991) (200) and others have reported that the hemin group of haemoglobin irreversibly binds to the *Taq* polymerase enzyme, inhibiting its reverse transcriptase activity during RT-PCR amplification (200-202).

4.2. NGS sequencing

Overall, we successfully amplified and sequenced 158 participant samples, which were sufficient to answer our research question. For the first time, our study compared the NGS sequencing results of two downstream analysis software, Deepchek® and Geneious®. The Deepchek®-HIV application is CE-IVD, which is specifically designed to perform genotypic analysis of HIV-1 drug resistance variants, whereas Geneious® is a general sequence analysis software, and requires creating a relevant workflow.

We observed high inconsistencies for DR variants reported between the Deepchek® and Geneious® analysis tools at proportions ≤ 4.5% which was subsequently used as our study cut-off. These inconsistencies could be due to PCR amplification errors and instrumental background noise.

Given the high sensitivity of NGS and its capacity to detect DRM as low as 1% of the quasispecies, it is more prone to detect PCR amplification errors as compared to Sanger sequencing (203). Sensitivity and accuracy of NGS has been shown to be dependent on the number of viral RNA templates that are extracted and amplified during RT-PCR. It is thus more especially prone to PCR amplification introduced artefacts and errors (204). PCR amplification has been associated with nucleotide misincorporation, observed mostly at viral proportions of 0.4-2% (203, 205-207).

Secondly, recombination between templates, mediated by PCR amplification has also been reported (208, 209). Experimental errors linked to PCR amplification have been observed mostly at frequency of 2% (205) to 5% (210). Dudley *et al., (*2017) investigated PCR mediated errors using viral clonal stocks from the HIV-1 HBX2 reference sequence. Sequence variations were considered as PCR amplification introduced errors. All variants reported were at ≤ 0.4% with only two variants occurring at a frequency of 0.9% and 0.6% and none of the variants were located at an HIV-1 DRM associated codon. Following resequencing, they found that mutations between 1% to 2% were inconsistently reported and thus they concluded that variants above the 2% threshold were authentic, and subsequently the used a 2% minimum threshold (205).

Additionally, discrepancies were also reported between the two analysis tools at ≥ 4.5%, with Geneious® unable to detect RT variants otherwise reported by Deepchek® (Table 3.4 and Appendix 3). The reason for such bioinformatic discrepancies needs to be further investigated. However it was outside the scope of expertise of the candidate/this study since it would involve looking at programme software changes

4.3. NGS vs Sanger sequencing

NGS been shown to detect more DRMs per HIV-1 patient sequence as compared to Sanger sequencing (170, 178, 184). The latter sequencing method is limited in DRM detection, as it is usually associated with a sensitivity of between 15 - 20%. In this study,

NGS reported more DRMs mainly because of the presence of DR minority variants occurring between 4.5% and 20%, as 14 minority HIV variants in 26 participants were detected in *PR* and 63 minority variants in 46 participants in RT *(*Appendix3).
In addition, Sanger failed to detect a total of four RT inhibitor mutations, M41I; Y181C; G190A and T215I presenting at proportions of between 20.0% and 25.5% of the viral population (Table.3.4) across three participants**.** This proved inconsistent given for example, the reporting of PI DRM, M41I at 21.0% in participant, KZC112 and M46I at 22.7% in participant MPC006.
These finding however agree with previous studies that reported inconsistencies of Sanger sequencing in detecting DRMs at 20%- 30% of the viral population (169, 184). Palmer *et al.*, (2005) analysed plasma samples from 26 patients experiencing VF on an nNRTI regimen and had prior exposure to at least two ARV drug classes. They used single-genome sequencing technique (SGS) and compared it to standard Sanger sequencing. All DRMs identified by Sanger sequencing were detected by SGS while Sanger failed to detect variants in 24 of the 26 patients as detected by SGS. Variants present at proportions of 10 to 35% by SGS were undetected 75% of the time by Sanger sequencing. Again, as expected, variants present in less than 10% of the viral population were almost never detected by Sanger sequencing (169).

Despite the differences in the numbers of detected DR mutations between both sequencing methods, high concordance of 96.8% for DRM ≥ 20% of the viral quasispecies was reported. Similar high concordance has been reported by other studies, including Tzou *et al.,* (2018) who reported 98.4% concordance using Sanger sequencing and Vela Diagnostic Sentosa® HIV genotyping assay methodologies, Le *et al.,* (2009), who compared Sanger sequencing and NGS platform Roche/454, reported 96.8% concordance, and lastly 100% was reported by Mohamed *et al.,* (2014), who compared Sanger sequencing and NGS on the Roche/454 platform.

NGS failed to detect majority PI mutations in two of the participants (Table 3.1). Since the Sanger and NGS sequences are phylogenetically related (Figure 3.4), the hypothesis for the observed discrepancy is possibly PCR resampling error due to low viral DNA input during PCR (211). It has been suggested that low target molecules in the original sample can result in certain viral strains being amplified more than others and resampled in the multiple rounds of PCR that occur. This leads to a skewed and incorrect population representation. A major limitation of this study is the lack of

availability of viral load measurements on the day of study sample collection for most of the participants, hence initial viral load and input DNA concentration prior to PCR cannot be accurately attained.

In an attempt to reduce PCR resampling bias, Boltz *et al*., 2015 (212) proposed implementing an RNA random primer ID tagging PCR amplification protocol. They showed a reduction in PCR amplification bias while increasing the number of starting cDNA template. Briefly, RNA is reverse transcribed using a specific cDNA primer attached to a 10-random-base ID tag with 22-mer primers containing uracils at the 5' end. This allows for improved adaptor and barcode ligation. This allows for reads with the same randomly generated IDs to be consolidated into a consensus sequence as representing a single cDNA species.

4.3.1. Protease inhibitor resistance

Different ARV drug resistance levels as defined by the Stanford HIVdb program correspond to different resistance score ranges. Susceptible (S) is defined by a drug resistance score between 0-10, potential low-level resistance (PLLR) is between 10-14, Low-level resistance (LLR) ranges from 15-30, intermediate resistance (IR) is 30-59 and high-level drug resistance (HLR) corresponds to a score > 60.

Treatment failure for patients on PI based regimens is characterised by the presence of PI DRMs, resulting in a cumulative genotypic resistance level of ≥15, corresponding to a phenotypic profile of LLR to HLR for the available PIs. Overall, NGS reported 12.7% participants harbouring HIV-1 quasispecies with PI DRM resulting in PI regimen failure compared to 11.4% by Sanger sequencing, with the difference not being statistically significant (p= 0.863).

The reported prevalence was lower than previously reported by other studies which reported a prevalence of 0% -26.6% using NGS (184-186). McKinnon *et al.,* (2012) used SGS to analyse plasma from 15 participants on a PI based regimen at viral rebound. They reported a prevalence of 26.6% for participants virologically failing with the presence of PI mutations (186). It is important to note that all these studies were limited by their small sample size which ranged from 7-36 participants. All participants included in the study had been on a PI regimen for a minimum of 48 weeks compared to 24

weeks in our study. Again, a 1% minimal threshold for DR variants detection was used, which is much lower than the ≥ 4.5 % used for our study.

If a 1% minimum threshold was used in this study, the prevalence of participants failing a PI- based regimen with the presence of PI DRM would be 20.3% (n= 32) compared to the current 12.7% (n=20) reported.

For the 12 additional participant sequences, the detection of PI DRMs at 1%-4.5%, resulted in PI resistance to at least two of the PIs included in this study. For example, the detection of PI DRMs, M46I (1.5%), I54V (1.7%), V82A (1.9%) and L90M (1.8%) in participant KZC124 would result in HLR to LPV/r and ATV/r while remaining susceptible to DRV/r. Additionally, in participant sequence KZC156, in which PI DRM Q58E was detected at 92.0% by Deepchek® but remained susceptible to all PIs, the additional detection of PI DRM V82F (1.0%) would result in intermediate resistance (IR) to LPV/r and LLR to both ATV/r and DRV/r. Thus, this participant would also be referred to third-line ART regimen. In participant sequence GPC014, in which the detection of L90M (94.0% Deepchek®), resulted in LLR to LPV/r and ATV/r (Table 3.3) and a subsequent referral to third-line ART regimen, the additional detection of I54L (1.0%), would result in LLR to DRV/r, thus the PI would no longer be a viable option as a third-line ART drug.

4.3.2. Protease inhibitor resistance at above 4.5%

PI DRMs were detected in 26 of the participants by NGS vs 24 participants by Sanger sequencing.
Of the 26 only 5 participants (19.0%) harboured PI minority variants and contrary to expectations, NGS only identified an additional two participants, MPC006 and GPC032 (20 vs 18 identified by Sanger sequencing), that harboured minority variants with PI resistance mutations resulting in inhibitor resistance and a subsequent referral to a third-line ART regimen.

Other publications have reported a range of prevalence rates of PI minority variants in patients at treatment failure, ranging from 0- 71.4% (154, 155, 185, 186). Le et al., (2009) reported the complete absence of major PI minority variants in 22 patients failing PI regimens. In another study, only two of the twelve participants (16.6%) harboured major minority PI variants (185). In a study by Fisher et al, (2012), prevalence of 71.4%

was reported amongst seven participants. The remainder studies (refs) had participant numbers ranging from 12- 22. It is important to note that these studies looked at a small number of participants, and our study is the highest to date.

Thus, 132 of the 158 participants (83.5%) failed a PI based regimen in the absence of known PR resistance mutations. Included in this group were 28 participants (21.2%) with no DRMs to any of the ARV drug classes. In the absence of ARV drug pressure, it is possible that some of the DR mutations that confer a fitness cost could have reverted to wild type virus. The time that the participants in this study were on a failing regimen is unclear, however, future studies looking at longitudinal cohorts could answer some of the questions raised about DR quasispecies diversity and evolution. Interestingly, there were no significant differences between the HIV viral load (p= 0.147), mean age (p= 0.833) and time on treatment (p=0.396) for the participants presenting with PI mutations and those without.

For participant GPC032, NGS detected four major PI HIVDR mutations (M46I, I54V, V82A and L90M) occurring at a viral population frequency of 4.5%-10% and hence were undetected by Sanger sequencing. The combined resistance scores of PI variants resulted in HLR to both ATV/r (score of 115) and LPV/r (score of 100), while the virus remained susceptible to DRV/r. M46I is selected primarily by all PIs except SQV and DRV. M46I/L has been observed to occur in about 20% and 10% of PI-treated patients, respectively. In our study we observed it in 8.9% of the participants. M46I usually occurs alone or in combination with V32I, I47V, L76V, I84V, and L90M while M46L usually occurs alone or in combination with I54V and V82. In participant GPC032, M46I occurred in combination with L90M. M46IL functions as both an inhibitory and compensatory mutation, as it is associated with reduced susceptibility but also increased PR catalytic efficiency to compensate for loss of viral fitness by other mutations**.** I54V which was also detected in 10.1% of the participant sequences is primarily selected for by IDV and LPV/. It contributes reduced susceptibility to each of the PIs except DRV. On its own it contributes a 15-resistance score to ATV/r and LPV/r each. V82A**,** also observed in 10.8% of the participant sequences**,** is a non-polymorphic mutation reducing susceptibility primarily to IDV and LPV (15 resistance score) while also contributing cross-resistance to each of the remaining PIs except DRV and TPV. Lastly, L90M (found in 1.3% of the participants) reduces susceptibility to each of the PIs except TPV and DRV, contributing a 25-resistance score.

In the other participant, MPC006, NGS detected major mutations, M46I in addition to K20T, V32I, I47V, I54V, V82A and L76V are detected also by Sanger sequencing. M46I alone results in potential low-level resistance to LPV/r and ATV/r as it has been shown to impact indirectly on the binding of inhibitor to the active site. It has also been shown to compensate for loss in viral fitness and viability due to other major mutations. L76V is selected by LPV and DRV. It reduces susceptibility to these PIs along with IDV and FPV (not available in the public sector in South Africa). Viral populations with L76V have been shown to have decreased drug susceptibility by 3.5 fold, while also severely reducing viral fitness (140, 213). The mutation is also compensatory, increasing susceptibility to ATV, SQV and TPV (214, 215). V32I and I47V were observed at a prevalence of 7.7% (n= 2) and 2.3% (n=4) in our cohort, respectively. Both are selected by PIs, IDV, FPV, LPV and DRV and reduce susceptibility to each of the PIs except SQV (138, 149, 216). V32I is usually in concert with I47 mutations, I47V or I47A. In combination with I47V it causes IR to both LPV and DRV and lays the groundwork for higher levels of resistance with additional DRMs. In combination with I47A it causes high-level LPV resistance and intermediate DRV resistance (217, 218). Lastly, K20T is a non-polymorphic PI-selected accessory mutation associated with reduced susceptibility to each of the PIs except DRV and TPV.

By current South African antiretroviral therapy (ART) guidelines, these two participants (MPC006 and GPC032) would be referred for a third-line ART regimen (Stanford genotypic resistant scores are ≥15 for both LPV/r and ATV/r) which comprises ofDRV/r, etravirine (ETR) and raltegravir (RAL) if NGS had been used to perform the genotypic test. These patients most likely remained on a failing PI based regimen, given that accredited Sanger sequencing protocols are used as the gold standard for HIV DR genotypic testing in diagnostics.

4.4. Participants with no change in PI resistance level

Interestingly, for three of the five participants with PI minority variants detected, there was no change in the overall genotypic resistance level**.**
In participant NWC016, minority variants M46L and V82S were a result of alternate amino acid changes to M46I and V82A already detected. M46L is a non-polymorphic mutation which does not contribute to resistance to any of the PIs while the V82S combination with V82A contributed a combined 30 resistance score to ATV/r, increasing the resistance score from 60 to 75. Despite this increase, there was no change in the genotypic resistance level as HLR is defined by a score of 60-75.

In participant KZC140, NGS detected major PI DRM M46I in addition to I54V. M46I which has been discussed above has resulted in an increased overall resistance score of 25. This increase in resistance score was not reflected in the genotypic resistance level, as participant sequence continued to confer LLR to both ATV/r and LPV/r.

Accessory mutation K20T was detected for participant NWC005 and has been shown to be a secondary accessory PI mutation frequent in treatment naïve patients but associated with reduced susceptibility to LPV/r and ATV/r. K20T only contributed a 5-point resistance score to ATV/r, with no major impact on the overall genotypic profile and thus participant NWC005 remained susceptible to all PIs.

4.5. Protease mutation pattern

The most prevalent PI drug resistant mutations among the 26 participants with detected PI DRMs were V82A, M46I, I54V and L76V. For participants with HLR (16/20) to either LPV/r or ATV/r, a combination of at least three of these PI mutations was observed. This speaks to the high genetic barrier to resistance to PR inhibitors, requiring at least three or four PI DRM to convey high level resistance (149, 150). This is unlike other drug classes like NRTIs, where the presence of one DRM is enough to confer high level resistance.

Notably a similar pattern/ combination of DRM was observed for participant GPC032, in which all DRMs were present as minority variants. This pattern of mutation emergence and ordered appearance has been extensively characterised for PI exposed participants

(143, 219). Watkins *et al.,* (2003) observed the accumulation of PI mutations in cell culture over increasing RTV drug concentrations. They observed the appearance of mutation V82A very early on, at concentrations as low as 0.34-0.198 µM of RTV. This was followed or was in concert with the emergence of mutation I54V. Mutation M46L was detected following round 12 of cell culture. At the highest drug concentration, 15µM, mutations at PR amino acid positions 10, 46, 54, 63, 82 and 84 were observed. This mutation combination was associated with a 1000-fold reduction in RTV drug susceptibility (219).

4.6. RT inhibitor minority variants

The presence of RT inhibitor minority variants has been shown to impact and predict failure on subsequent RT inhibitor containing regimens. NGS detected RT inhibitor resistance in 130 of the 158 participants, seven more than were detected by Sanger sequencing. For the 29 participants with a resultant change in the level of genotypic resistance to the selected RT inhibitors, there was new and higher resistance to ABC, FTC, AZT, TDF and EFV for 12, 6, 14, 6, and 13 of the participants, respectively (Table 3.4).

Most interesting of the 29 participants only three also harboured virus sequences with PI resistance and would be recommended for a salvage integrase third- line regimen. (Table.3.4). For participant ECC018, the additional detection of DR RT minority variants resulted in IR to TDF and AZT, but virus was still susceptible to ETR. Participant ECC018 also had high level resistance to LPV/r and ATV/r but LLR to DRV/r. Currently the SA third- line committee makes use of an algorithm to streamline regimen selection for patients. All patients will get DRV/r and 3TC or FTC plus either TDF or AZT (whichever has the lowest resistance score). If the TDF or AZT score is 30-59 (IR) or if the DRV score is ≥15, RAL is added. If the TDF and AZT score is >29 and DRV/r score ≥15, ETR in addition to RAL is added (unless ETR score is >29) (97). By use of the algorithm rules (97), participant ECC018 would be referred and changed to a third-line regimen consisting of DRV/r, 3TC/FTC with ETR and RAL. Should the third line selection be solely based on Sanger sequencing results, the participant would be prescribed TDF (a lower resistant score), 3TC/FTC with ETR and RAL.

In participant MPC006, NGS sequencing results reported HLR to LPV/r ATV/r and DRV/r, LLR to TDF and intermediate resistance to AZT in combination with LLR to ETR. For this participant, TDF, 3TC/FTC and DRV/r, RAL would be possibly prescribed as a third-line regimen. If regimen selection was based on Sanger results only, ETR would not be included as the participant would be reported as having low-level resistance to TDF. In participant KZC140 a similar regimen would be prescribed (DRV/r- 3TC/FTC) regardless of the sequencing technique used, except for the addition of TDF as opposed to AZT as part of the RT inhibitor backbone as low-level resistance was reported following detection of minority DR variants by NGS.

4.7. RT inhibitor mutations

The most prevalent RT minority variants were D67E, K103N/S and M184V/I.

4.7.1. D67E

D67E (n-=28, 17.7%) is a non-polymorphic TAM conferring LLR to NRTIs, AZT and d4T. When present with other TAMs, it contributes reduced susceptibility to other NRTIs such as ABC, AZT, and TDF. For example, in participant MPC006, the detection of D67E in combination with other TAMs, K70R and K219E resulted in increased levels of resistance for RT inhibitors ABC (HLR), TDF (LLR) and AZT (IR) (Table 3.4).

4.7.2. K103N

The most prevalent minority nNRTI mutation in the cohort was K103N (n=16, 8.9%). K103N is a non-polymorphic mutation selected and conferring high-level resistance to NVP and EFV by approximately 50 and 20-fold, respectively. It is also associated with a modest reduction in viral fitness (220). Detection of K103N especially as minority variants is most likely representative of an nNRTI regimen history, as most participants have been previously exposed to EFV as part of their first-line nNRTI based regimen.

EFV has a long half-life, allowing for K103N and other mutations such as V106M, G190A selected for primarily by the inhibitor to also persist as part of the viral quasispecies. For example, in participant sequence KZC183, minority variants, K103N, E138G, G190A and F227L were detected (Table 3.4). This participant had been on a first-line regimen consisting of d4T-3TC-EFV and following failure of the nNRTI regimen

had been switched from the regimen for at least 13 months. nNRTI mutations such as K103N and G190A have been shown to persist for up to 3.7- 5 years, in the absence of EFV drug pressure even after short term exposures at first-line treatment (221).

4.7.3. M184V/I

The most frequent minority NRTI mutation was M184V/I (n=13, 3.8%). The mutation is selected for by 3TC/FTC and confers HLR to 3TC/FTC, emerging as early as within 6 weeks on the inhibitor. It was surprising that overall most patients did not harbour M184V/I (n=68) and it was detected as a minority variant in some participants (n=13) (Figure 3.4) since it should continue to be selected for under a PI based regimen that includes 3TC. Future analysis of samples from a longitudinal cohort could provide answers. In participant KZC112 (Table 3.4) who had been exposed to a 3TC containing PI based regimen for at least 14 months (Appendix 2), M184V was detected as a minority variant (Deepchek®, 20.9%) along with nNRTI mutation K103N (Deepchek®, 15.2%). This could signal reversion of the mutation to wild-type, which has been shown to occur approximately one year following drug pressure removal probably because of its high cost on viral fitness (126).

When the mutation is present alone, it can cause hyper-susceptibility to AZT, d4T, and TDF but when detected along with TAMs or K65R, it may partially reverse the loss of susceptibility caused by those mutations**.** In participant sequenceKZC133, M184V/I was detected along with K65R. K65R confers IR to ABC and HLR to TDF. But the additional detection of M184V/I resulted in LLR to ABC and S to TDF (Table 3.4).

4.8. Future work

Most patients failing a PI based regimen harbour no PI drug resistant mutations. Evidence has also accumulated to suggest that mutations located on other amino acid sites outside of PR, most notably within the Gag cleavage sites, can contribute to PI drug resistance(60, 140). Unfortunately, these were not evaluated as part of this study. Future work should investigate the interaction of these factors in a single cohort. Thus, the next steps in this study will be to investigate these additional factors in the 158 participants.

4.8.1. Adherence

It has been shown that non-adherence is a driver for observed virological failure on PI based regimens (153, 156, 158) without the presence of PI DRMs. Walsh *et al.,* (2006) investigated the association of drug adherence, virological failure and the presence of DR mutations in a cohort of patients on PI regimens. They compared this association between patients on their regimens for ≥ 6 months with a viral load of ≥ 1000 RNA copies/ml and those with a viral of ≤ 50 RNA copies/ml. Adherence was measured using an electric measuring system. In the viremic group (VL≥ 1000 RNA copies/ml), patients with full PI susceptibility with no PI DRMs had an adherence of 63.3%, those with intermediate resistance showed 85.1% adherence and high-level resistance with major PI DRMs was linked to an adherence of 90.8%. They concluded that the number of DR mutations increases linearly with improving adherence in viremic patients (222). Findings were supported by Bangberg *et al.,* (2005), as they also found that PR inhibitor resistance was limited to patients with 65% to 100% adherence (223). Van Zyl *et al.,* (2009), found LPV plasma concentration of ≥1mg/mL and LPV hair concentration of ≥3.63 ng/mg as associated with virologic failure of 92% and 96%, respectively (153).

Most of these studies used patient self-reporting and pill counting as measures of adherence. However, this method is highly unreliable as most patients give false reports and can adjust the pill count to favour adherence.
Future work should focus on conducting adherence tests on available plasma, thus eliminating or confirming non-adherence as a cause for observed virological failures with no DRM. We hypothesis that, LPV/r plasma levels will be low, for the majority of these participants more especially in the 28 participants with no DRM to any ARV drugs.

4.8.2. Gag mutations

Evidence describing and pointing to co-evolution of mutations in the Gag (p55) cleavage sites, which is a natural substrate of PR is available. Mutations in Gag develop with the aim of compensating for a loss in viral fitness due to the development of PI resistant mutations in drug exposed patients (60, 140, 185, 224, 225). There is now however accumulating evidence that mutations developing in *gag* in patients exposed to PIs also directly confer and contribute to PI resistance in the presence and absence of mutations in PR(160, 163-166, 226, 227).

Mutations located in the Gag cleavage sites, including P7/PI and P1/P6 and those outside the cleavage site have been studied. Dam *et al.,* (2009) investigated the effects of PR mutations and Gag mutations against six distinct PIs: IDV, NFV, APV, SQV, LPV and ATV. They used 4 clones from 6 patients. Patient-derived PR, RT and complete *gag* sequences were found to be strongly resistant to most of the PIs. Also clones that carried patient-derived *gag* but wild-type PR and RT, exhibited increased IC_{50} levels ranging from 1.6-fold to 5.6-fold, relative to the reference isolate NL4.3, These findings suggested that even in the presence of wild-type PR, changes in *gag* occurring during under selective pressure by PIs has a clear effect on HIV-1 susceptibility (164). In contrast, resistance was found markedly lower in viruses carrying only patient-derived PR and RT in the absence of any patient-derived *gag* sequences. Mutations in *gag* associated with these observed reductions in susceptibility, A431 and I437V were identified in the Gag NC-SP2-P6 region. This finding emphasizes the critical importance of HIV-1 Gag cleavage site mutations in contributing to viral resistance to PIs.

All these studies have led to the hypothesis that for some of the 138 participants in our cohort failing treatment without associated PI drug resistant mutation, may harbour mutations in the Gag cleavage sites. The presence of these mutations could explain virological failure in these participants, thus future work should focus on sequencing and characterizing mutations in full length Gag for this cohort.

4.9. Conclusion

Overall the presence of PI DRMs corresponding to PI based treatment failure was detected in only 12.7% of the cohort. Of these, the use of NGS identified only an additional two participants who harboured HIV-1 minority variants with PI resistance that would warrant they be switched to a third-line regimen. The use of NGS also proved beneficial for some of the participants for which RT inhibitor minority variants were detected. The presence of RT inhibitor DR minority variants correlated with the drug history of the participants, as the majority were on an AZT/TDF-3TC-EFV first-line regimen (Appendix 2) which selects for variants K103NS, D67E and E138AG, which were the most prevalent RT mutations identified. For each of these participants the resultant change in the RT inhibitor genotypic and phenotypic profiles would impact on the choice of the third-line RT regimen backbone. There is no current consensus on the

limit of detection for which detected minority variants are to be considered clinically relevant. Evidence pointing to the significance of minority nNRTI variants in treatment naïve and experienced patients suggests a threshold of 1%(179). But in the case of PI drug mutations, this threshold has not been proven to be significant. This suggests that the detection threshold for clinical relevance may be drug class specific. Factors to be considered when selecting a threshold should include the genetic barrier of the drug, the interaction of the mutations, clinical characteristics of the patient, most notably the viral load, and levels of adherence. For example, M46I/L is known to not only contribute to resistance to LPV/r, but it also has compensatory effects on viral fitness. If this mutation was to be detected in a patient with a viral load of 10 000 RNA copies/ml, then the mutational load would be 200 RNA copies/ml. It has been suggested that a mutational load of ≥ 2000 RNA copies/ml is significant. This all combined with low LPV/r plasma concentrations would render the mutation at that proportion clinically irrelevant.

Adherence counselling tactics and frequency may also need to be increased as for the majority of the patients, non-adherence may be an issue. This is supported by 28 participants in our study who had no detectable ARV drug resistant mutations associated with any drug class. Again, participants would be expected to harbour 3TC/FTC associated mutation, M184I/V, which is commonly selected for in patients prescribed any of the two ARV drugs but was only detected in 68 participant sequences. M184I/V has been shown to rapidly revert to wild type following removal of drug pressure, which would be the case for regimen non-adherent participants.

In conclusion, this study reports on HIV-1 DR sequence data from the largest numbers of individuals failing a PI based regimen and the detection of PI minority variants using NGS. Findings of this study highlight the infrequency at which PI minority variants at any proportion are detected, corroborating previous findings stating that most patients failing on a PI based regimen harbour no PR drug resistant associated mutations. This study has shown no significant (p=0863) additional benefits of using NGS over Sanger sequencing in diagnostics to identify patients with potential PI mutations needing to be switched to a third line regimen. However, genotypic testing by NGS could impact on the optimal third-line regimen selection, especially in selecting an RT backbone. This could prove mostly beneficial for patients without a known full ARV treatment history, thus allowing for better clinical outcomes of salvage therapy.

5. References

1. Friedman-Kien A, Laubenstein L, Marmor M, Hymes K, Green J, Ragaz A, et al. Kaposis sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. MMWR Morbidity and mortality weekly report. 1981;30(25):305-8.
2. Barre F, Chermann J, Rey F, Nugeyre M, Chamaret S, Gruest J, et al. Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum and L. Montagnier, Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science. 1983;220:868-71.
3. unaids. fact sheet2016.
4. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. The early spread and epidemic ignition of HIV-1 in human populations. Science. 2014;346(6205):56-61.
5. Marx JL. New disease baffles medical community. Science. 1982;217(4560):618-21.
6. Control CfD. A cluster of Kaposi's sarcoma and Pneumocystis carinii pneumonia among homosexual male residents of Los Angeles and Orange Counties, California. MMWR Morbidity and mortality weekly report. 1982;31(23):305.
7. Jaffe HW, Choi K, Thomas PA, Haverkos HW, Auerbach DM, Guinan ME, et al. National case-control study of Kaposi's sarcoma and Pneumocystis carinii pneumonia in homosexual men: part 1, epidemiologic results. Annals of internal medicine. 1983;99(2):145-51.
8. Polsky B, Gold JW, Whimbey E, Dryjanski J, Brown AE, Schiffman G, et al. Bacterial pneumonia in patients with the acquired immunodeficiency syndrome. Annals of internal medicine. 1986;104(1):38-41.
9. Colombini EG, Oroszlan S, Staal S, Gallo R. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). 1983.
10. Hahn BH, Shaw GM, Arya SK, Popovic M, Gallo RC, Wong-Staal F. Molecular cloning and characterization of the HTLV-III virus associated with AIDS. Nature. 1984;312(5990):166-9.
11. Peeters M, Courgnaud V, Abela B, Auzel P, Pourrut X, Bibollet-Ruche F, et al. Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. Emerg Infect Dis [serial on the Internet]. 2002.
12. Wolfe ND, Switzer WM, Carr JK, Bhullar VB, Shanmugam V, Tamoufe U, et al. Naturally acquired simian retrovirus infections in central African hunters. The Lancet. 2004;363(9413):932-7.
13. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, et al. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature. 1999;397(6718):436-41.
14. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR. An African primate lentivirus (SIVsmclosely related to HIV-2. 1989.
15. Horsburgh CR, Holmberg SD. The global distribution of human immunodeficiency virus type 2 (HIV-2) infection. Transfusion. 1988;28(2):192-5.
16. Fanales-Belasio E, Raimondo M, Suligoi B, Buttò S. HIV virology and pathogenetic mechanisms of infection: a brief overview. Annali dell'Istituto superiore di sanita. 2010;46:5-14.
17. Ariën KK, Abraha A, Quinones-Mateu ME, Kestens L, Vanham G, Arts EJ. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. Journal of virology. 2005;79(14):8979-90.
18. Marlink R, Kanki P, Thior I, Travers K, Eisen G, Siby T, et al. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. Science. 1994;265(5178):1587-91.
19. De Cock KM, Adjorlolo G, Ekpini E, Sibailly T, Kouadio J, Maran M, et al. Epidemiology and transmission of HIV-2: why there is no HIV-2 pandemic. Jama. 1993;270(17):2083-6.
20. de Silva TI, Cotten M, Rowland-Jones SL. HIV-2: the forgotten AIDS virus. Trends in microbiology. 2008;16(12):588-95.
21. Clavel F. Isolation of a new human retrovirus from West African patients with AIDS. Science. 1986;233:343-7.
22. Damond F, Apetrei C, Robertson DL, Souquière S, Leprêtre A, Matheron S, et al. Variability of human immunodeficiency virus type 2 (HIV-2) infecting patients living in France. Virology. 2001;280(1):19-30.

23.da Silva ZJ, Oliveira I, Andersen A, Dias F, Rodrigues A, Holmgren B, et al. Changes in prevalence and incidence of HIV-1, HIV-2 and dual infections in urban areas of Bissau, Guinea-Bissau: is HIV-2 disappearing? Aids. 2008;22(10):1195-202.
24.van Tienen C, van der Loeff MS, Zaman SM, Vincent T, Sarge-Njie R, Peterson I, et al. Two distinct epidemics: the rise of HIV-1 and decline of HIV-2 infection between 1990 and 2007 in rural Guinea-Bissau. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2010;53(5):640-7.
25.Simon F, Mauclère P, Roques P, Loussert-Ajaka I, Müller-Trutwin MC, Saragosti S, et al. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nature medicine. 1998;4(9).
26.Gürtler LG, Hauser PH, Eberle J, Von Brunn A, Knapp S, Zekeng L, et al. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. Journal of virology. 1994;68(3):1581-5.
27.Triques K, Bourgeois A, Vidal N, Mpoudi-Ngole E, Mulanga-Kabeya C, Nzilambi N, et al. Near-full-length genome sequencing of divergent African HIV type 1 subtype F viruses leads to the identification of a new HIV type 1 subtype designated K. AIDS research and human retroviruses. 2000;16(2):139-51.
28.Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature. 2008;455(7213):661-4.
29.Robertson D, Anderson J, Bradac J, Carr J, Foley B, Funkhouser R, et al. HIV-1 nomenclature proposal. Science. 2000;288(5463):55-.
30.Myers G, MacInnes K, KORBER B. The emergence of simian/human immunodeficiency viruses. AIDS research and human retroviruses. 1992;8(3):373-86.
31.Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The challenge of HIV-1 subtype diversity. New England Journal of Medicine. 2008;358(15):1590-602.
32.gov l. HIV Circulating Recombinant Forms (CRFs): lanl.gov; 2017 [Available from: https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html.
33.Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, et al. Timing the ancestor of the HIV-1 pandemic strains. Science. 2000;288(5472):1789-96.
34.Bell NM, Lever AM. HIV Gag polyprotein: processing and early viral particle assembly. Trends in microbiology. 2013;21(3):136-44.
35.Suzuki Y, Suzuki Y. Gene regulatable lentiviral vector system. Viral gene therapy: IntechOpen; 2011.
36.Turner BG, Summers MF. Structural biology of HIV. Journal of molecular biology. 1999;285(1):1-32.
37.Gelderblom H, Özel M, Pauli G. Morphogenesis and morphology of HIV structure-function relations. Archives of virology. 1989;106(1-2):1-13.
38.Mandell GL, Bennett JE, Dolin R, Schwartz DA. Principles and practice of infectious disease. Archives of Pathology and Laboratory Medicine. 1997;121(8):908.
39.Laskey SB, Siliciano RF. A mechanistic theory to explain the efficacy of antiretroviral therapy. Nature Reviews Microbiology. 2014;12(11):772.
40.Geijtenbeek TB, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GC, Middel J, et al. DC-SIGN, a dendritic cell–specific HIV-1-binding protein that enhances trans-infection of T cells. Cell. 2000;100(5):587-97.
41.Saphire AC, Bobardt MD, Zhang Z, David G, Gallay PA. Syndecans serve as attachment receptors for human immunodeficiency virus type 1 on macrophages. Journal of virology. 2001;75(19):9187-200.
42.Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature. 1984;312(5996):763-7.
43.McDougal J, Kennedy M, Sligh J, Cort S, Mawle A, Nicholson J. Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. Science. 1986;231(4736):382-5.

44.Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature. 1998;393(6686):648.
45.Chen B, Vogan EM, Gong H, Skehel JJ, Wiley DC, Harrison SC. Structure of an unliganded simian immunodeficiency virus gp120 core. Nature. 2005;433(7028):834.
46.Berger EA, Doms R, Fenyö E-M, Korber B, Littman D, Moore J, et al. A new classification for HIV-1. Nature. 1998;391:240.
47.Alkhatib G, Combadiere C, Broder CC, Feng Y. CC CKR5: A RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-topic HIV-1. Science. 1996;272(5270):1955.
48.Doms RW, Peiper SC. Unwelcomed guests with master keys: how HIV uses chemokine receptors for cellular entry. Virology. 1997;235(2):179-90.
49.Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annual review of immunology. 1999;17(1):657-700.
50.Weissenhorn W, Dessen A, Harrison S, Skehel J, Wiley D. Atomic structure of the ectodomain from HIV-1 gp41. Nature. 1997;387(6631):426.
51.Chan DC, Fass D, Berger JM, Kim PS. Core structure of gp41 from the HIV envelope glycoprotein. Cell. 1997;89(2):263-73.
52.Melikyan GB. Common principles and intermediates of viral protein-mediated fusion: the HIV-1 paradigm. Retrovirology. 2008;5(1):111.
53.Arhel N. Revisiting HIV-1 uncoating. Retrovirology. 2010;7(1):96.
54.Stein BS, Gowda SD, Lifson JD, Penhallow RC, Bensch KG, Engleman EG. pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. Cell. 1987;49(5):659-68.
55.Goff SP. Retroviral reverse transcriptase: synthesis, structure, and function. Journal of acquired immune deficiency syndromes. 1990;3(8):817-31.
56.Hu W-S, Hughes SH. HIV-1 reverse transcription. Cold Spring Harbor perspectives in medicine. 2012;2(10):a006882.
57.Peliska JA, Benkovic SJ. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Science. 1992;258(5085):1112-8.
58.Sousa R, Chung YJ, Rose JP, Wang B-C. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. Nature. 1993;364(6438):593-9.
59.Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. Journal of virology. 2010;84(19):9864-78.
60.Clavel F, Hance AJ. HIV drug resistance. New England Journal of Medicine. 2004;350(10):1023-35.
61.Ellison V, Brown PO. A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro. Proceedings of the National Academy of Sciences. 1994;91(15):7316-20.
62.Colicelli J, Goff SP. Sequence and spacing requirements of a retrovirus integration site. Journal of molecular biology. 1988;199(1):47-59.
63.Vink C, Lutzke RAP, Plasterk RH. Formation of a stable complex between the human immunodeficiency virus integrase protein and viral DNA. Nucleic acids research. 1994;22(20):4103-10.
64.Roe T, Chow SA, Brown PO. 3'-end processing and kinetics of 5'-end joining during retroviral integration in vivo. Journal of virology. 1997;71(2):1334-40.
65.Bukrinsky MI, Haggerty S, Dempsey MP, Sharova N, Adzhubei A, Spitz L, et al. A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. Nature. 1993;365(6447):666.
66.Gallay P, Hope T, Chin D, Trono D. HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. Proceedings of the National Academy of Sciences. 1997;94(18):9825-30.

67.Heinzinger NK, Bukinsky M, Haggerty SA, Ragland AM, Kewalramani V, Lee M-A, et al. The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. Proceedings of the National Academy of Sciences. 1994;91(15):7311-5.
68.Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, et al. A role for LEDGF/p75 in targeting HIV DNA integration. Nature medicine. 2005;11(12):1287.
69.Coffin JM, Hughes SH, Varmus HE. The interactions of retroviruses and their hosts: Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY); 1997.
70.Wiskerchen M, Muesing MA. Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells. Journal of virology. 1995;69(1):376-86.
71.Keele BF, Derdeyn CA. Genetic and antigenic features of the transmitted virus. Current Opinion in HIV and AIDS. 2009;4(5):352-7.
72.Eigen M, Schuster P. A principle of natural self-organization. Naturwissenschaften. 1977;64(11):541-65.
73.Eigen M, Schuster P. The hypercycle: a principle of natural self-organization: Springer Science & Business Media; 2012.
74.Andino R, Domingo E. Viral quasispecies. Virology. 2015;479:46-51.
75.Page KM, Nowak MA. Unifying evolutionary dynamics. Journal of theoretical biology. 2002;219(1):93-8.
76.Kahn JO, Walker BD. Acute human immunodeficiency virus type 1 infection. New England Journal of Medicine. 1998;339(1):33-9.
77.Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. Aids. 2003;17(13):1871-9.
78.Mattapallil JJ, Douek DC, Hill B, Nishimura Y, Martin M, Roederer M. Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. Nature. 2005;434(7037):1093.
79.Bollinger RC, Brookmeyer RS, Mehendale SM, Paranjape RS, Shepherd ME, Gadkari DA, et al. Risk factors and clinical presentation of acute primary HIV infection in India. Jama. 1997;278(23):2085-9.
80.Rosenberg ES, Billingsley JM, Caliendo AM, Boswell SL, Sax PE, Kalams SA, et al. Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia. Science. 1997;278(5342):1447-50.
81.Mackewicz E, Levy J, Yang L, Lifson J. Non-cytolytic CD8 T-cell anti-HIV responses in primary HIV-1 infection. The Lancet. 1994;344(8938):1671-3.
82.Gomez A, Smaill F, Rosenthal K. Inhibition of HIV replication by CD8+ T cells correlates with CD4 counts and clinical stage of disease. Clinical & Experimental Immunology. 1994;97(1):68-75.
83.Walker CM, Moody DJ, Stites DP, Levy JA. CD8 lymphocytes can control HIV infection in vitro by suppressing virus replication. Science. 1986;234:1563-7.
84.Naif HM. Pathogenesis of HIV infection. Infectious disease reports. 2013;5(Suppl 1).
85.Alizon S, von Wyl V, Stadler T, Kouyos RD, Yerly S, Hirschel B, et al. Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. PLoS pathogens. 2010;6(9):e1001123.
86.Mellors JW, Rinaldo Jr CR, Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. SCIENCE-NEW YORK THEN WASHINGTON-. 1996:1167-9.
87.Redfield RR, Burke DS. HIV Infection. Scientific American. 1988;259(4):90-9.
88.Palella Jr FJ, Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. New England Journal of Medicine. 1998;338(13):853-60.
89.Organisation Wh. WHO Drug Resistance Report 2017. 2018.
90.Group ISS. Initiation of antiretroviral therapy in early asymptomatic HIV infection. N Engl J Med. 2015;2015(373):795-807.

91.Rodger A, Bruun T, Cambiano V, Vernazza P, Strada V, Van Lunzen J, editors. 153LB: HIV transmission risk through condomless sex if HIV+ partner on suppressive ART: PARTNER Study. 21st Conference on Retroviruses and Oppotunistic Infections; 2014.
92.Donnell D, Baeten JM, Kiarie J, Thomas KK, Stevens W, Cohen CR, et al. Heterosexual HIV-1 transmission after initiation of antiretroviral therapy: a prospective cohort analysis. The Lancet. 2010;375(9731):2092-8.
93.south african national hiv treatment guidelines 2013
In: department sanh, editor. 2013.
94.WHO-CDS. UPDATED RECOMMENDATIONS ON FIRST-LINE AND SECOND-LINE ANTIRETROVIRAL REGIMENS AND POST-EXPOSURE PROPHYLAXIS AND RECOMMENDATIONS ON EARLY INFANT DIAGNOSIS OF HIV https://sahivsoc.org/Files/WHO-CDS-HIV-18.18-eng.pdfJULY, 2018.
95.Thorne C, Newell M-L. Prevention of mother-to-child transmission of HIV infection. Current opinion in infectious diseases. 2004;17(3):247-52.
96.Meintjes G, Moorhouse MA, Carmona S, Davies N, Dlamini S, Van Vuuren C, et al. Adult antiretroviral therapy guidelines 2017. Southern African journal of HIV medicine. 2017;18(1).
97.M Moorhouse FC, G Maartens, WDF Venter, M Moosa, K Jamaloodien, and MP Fox. An Algorithm to Determine Appropriate Third-line Antiretroviral Regimens in Patients Failing Protease Inhibitor-based Regimens in Resource-limited Settings. . 2017.
98.Chun T-W, Davey Jr RT, Engel D, Lane HC, Fauci AS. AIDS: re-emergence of HIV after stopping therapy. Nature. 1999;401(6756):874.
99.Hart G, Orr D, Penn C, Figueiredo H, Gray N, Boehme R, et al. Effects of (-)-2'-deoxy-3'-thiacytidine (3TC) 5'-triphosphate on human immunodeficiency virus reverse transcriptase and mammalian DNA polymerases alpha, beta, and gamma. Antimicrobial agents and chemotherapy. 1992;36(8):1688-94.
100. Stein DS, Moore KH. Phosphorylation of nucleoside analog antiretrovirals: a review for clinicians. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy. 2001;21(1):11-34.
101. Furman PA, Fyfe JA, St Clair MH, Weinhold K, Rideout JL, Freeman GA, et al. Phosphorylation of 3'-azido-3'-deoxythymidine and selective interaction of the 5'-triphosphate with human immunodeficiency virus reverse transcriptase. Proceedings of the National Academy of Sciences. 1986;83(21):8333-7.
102. Götte M. Inhibition of HIV-1 reverse transcription: basic principles of drug action and resistance. Expert review of anti-infective therapy. 2004;2(5):707-16.
103. Kati WM, Johnson KA, Jerva L, Anderson K. Mechanism and fidelity of HIV reverse transcriptase. Journal of Biological Chemistry. 1992;267(36):25988-97.
104. Huang H, Chopra R, Verdine GL, Harrison SC. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science. 1998;282(5394):1669-75.
105. Goody R, Müller B, Restle T. Factors contributing to the inhibition of HIV reverse transcriptase by chain-terminating nucleotides in vitro and in vivo. FEBS letters. 1991;291(1):1-5.
106. Sarafianos SG, Marchand B, Das K, Himmel DM, Parniak MA, Hughes SH, et al. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. Journal of molecular biology. 2009;385(3):693-713.
107. St Clair MH, Richards C, Spector T, Weinhold K, Miller W, Langlois AJ, et al. 3'-Azido-3'-deoxythymidine triphosphate as an inhibitor and substrate of purified human immunodeficiency virus reverse transcriptase. Antimicrobial agents and chemotherapy. 1987;31(12):1972-7.
108. König H, Behr E, Löwer J, Kurth R. Azidothymidine triphosphate is an inhibitor of both human immunodeficiency virus type 1 reverse transcriptase and DNA polymerase gamma. Antimicrobial agents and chemotherapy. 1989;33(12):2109-14.
109. Vrang L, Oberg B, Löwer J, Kurth R. Reverse transcriptases from human immunodeficiency virus type 1 (HIV-1), HIV-2, and simian immunodeficiency virus (SIVMAC) are susceptible to inhibition by foscarnet and 3'-azido-3'-deoxythymidine triphosphate. Antimicrobial agents and chemotherapy. 1988;32(11):1733-4.

110. Cihlar T, Ray AS. Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. Antiviral research. 2010;85(1):39-58.
111. Martin JC, Hitchcock MJ, De Clercq E, Prusoff WH. Early nucleoside reverse transcriptase inhibitors for the treatment of HIV: a brief history of stavudine (D4T) and its comparison with other dideoxynucleosides. Antiviral research. 2010;85(1):34-8.
112. Cossarizza A, Moyle G. Antiretroviral nucleoside and nucleotide analogues and mitochondria. Aids. 2004;18(2):137-51.
113. Baba M, De Clercq E, Tanaka H, Ubasawa M, Takashima H, Sekiya K, et al. Potent and selective inhibition of human immunodeficiency virus type 1 (HIV-1) by 5-ethyl-6-phenylthiouracil derivatives through their interaction with the HIV-1 reverse transcriptase. Proceedings of the National Academy of Sciences. 1991;88(6):2356-60.
114. Balzarini J. Current status of the non-nucleoside reverse transcriptase inhibitors of human immunodeficiency virus type 1. Current topics in medicinal chemistry. 2004;4(9):921-44.
115. Jochmans D, Vingerhoets J, Arnoult E, Geeraert L, Guillemont J. Human immunodeficiency virus type 1 non-nucleoside reverse transcriptase inhibitors. Antiviral research: American Society of Microbiology; 2009. p. 33-50.
116. Kohlstaedt L, Wang J, Friedman J, Rice P, Steitz T. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science. 1992;256(5065):1783-90.
117. Hsiou Y, Ding J, Das K, Clark Jr A, Hughes S, Arnold E. Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: implications of conformational changes for polymerization and inhibition mechanisms. Structure. 1996;4(7):853-60.
118. Spence RA, Kati WM, Anderson KS, Johnson KA. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. SCIENCE-NEW YORK THEN WASHINGTON-. 1995:988-.
119. D'cruz OJ, Uckun FM. Dawn of non-nucleoside inhibitor-based anti-HIV microbicides. Journal of Antimicrobial Chemotherapy. 2006;57(3):411-23.
120. Sluis-Cremer N, Temiz NA, Bahar I. Conformational changes in HIV-1 reverse transcriptase induced by nonnucleoside reverse transcriptase inhibitor binding. Current HIV Research. 2004;2(4):323-32.
121. de Béthune M-P. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989–2009). Antiviral research. 2010;85(1):75-90.
122. Rodgers D, Gamblin S, Harris B, Ray S, Culp J, Hellmig B, et al. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. Proceedings of the National Academy of Sciences. 1995;92(4):1222-6.
123. Balzarini J, De Clercq E. [25] Analysis of inhibition of retroviral reverse transcriptase. Methods in enzymology. 275: Elsevier; 1996. p. 472-502.
124. Roberts NA, Martin JA, Kinchington D, Broadhurst AV, Craig JC, Duncan IB, et al. Rational design of peptide-based HIV proteinase inhibitors. Science. 1990;248(4953):358-61.
125. Wensing AM, Calvez V, Günthard HF, Johnson VA, Paredes R, Pillay D, et al. 2017 update of the drug resistance mutations in HIV-1. Topics in antiviral medicine. 2016;24(4):132.
126. Castro H, Pillay D, Cane P, Asboe D, Cambiano V, Phillips A, et al. Persistence of HIV-1 transmitted drug resistance mutations. The Journal of infectious diseases. 2013;208(9):1459-63.
127. Rhee S-Y, Jordan MR, Raizes E, Chua A, Parkin N, Kantor R, et al. HIV-1 drug resistance mutations: potential applications for point-of-care genotypic resistance testing. PLoS ONE. 2015;10(12):e0145772.
128. Vici Varghese RS, Rhee S-Y, Liu T, Simen BB, Egholm M, Hanczaruk B, et al. Minority Variants Associated with Transmitted and Acquired HIV-1 Non-Nucleoside RT Inhibitor (NNRTI) Resistance: Implications for the Use of Second Generation NNRTIs. Journal of acquired immune deficiency syndromes (1999). 2009;52(3):309.

129. World Health Organization USCfDCaP, The Global Fund to Fight AIDS, Tuberculosis and Malaria. HIV drug resistance report 2017. https://www.who.int/hiv/pub/drugresistance/hivdr-report-2017/en/2017.
130. Naeger LK, Margot NA, Miller MD. ATP-dependent removal of nucleoside reverse transcriptase inhibitors by human immunodeficiency virus type 1 reverse transcriptase. Antimicrobial agents and chemotherapy. 2002;46(7):2179-84.
131. Basavapathruni A, Bailey CM, Anderson KS. Defining a molecular mechanism of synergy between nucleoside and nonnucleoside AIDS drugs. Journal of Biological Chemistry. 2004;279(8):6221-4.
132. Martin M, Del Cacho E, Codina C, Tuset M, De Lazzari E, Mallolas J, et al. Relationship between adherence level, type of the antiretroviral regimen, and plasma HIV type 1 RNA viral load: a prospective cohort study. AIDS research and human retroviruses. 2008;24(10):1263-8.
133. Quiñones-Mateu ME, Arts EJ. HIV-1 fitness: implications for drug resistance, disease progression, and global epidemic evolution. HIV sequence compendium. 2001;2001:134-70.
134. Ali A, Bandaranayake RM, Cai Y, King NM, Kolli M, Mittal S, et al. Molecular basis for drug resistance in HIV-1 protease. Viruses. 2010;2(11):2509-35.
135. Kuiper BD, Keusch BJ, Dewdney TG, Chordia P, Ross K, Brunzelle JS, et al. The L33F darunavir resistance mutation acts as a molecular anchor reducing the flexibility of the HIV-1 protease 30s and 80s loops. Biochemistry and biophysics reports. 2015;2:160-5.
136. Agniswamy J, Shen C-H, Aniana A, Sayer JM, Louis JM, Weber IT. HIV-1 protease with 20 mutations exhibits extreme resistance to clinical inhibitors through coordinated structural rearrangements. Biochemistry. 2012;51(13):2819-28.
137. Prabu-Jeyabalan M, King N, Nalivaika E, Scott W, Schiffer C, editors. Drug resistance and substrate recognition in HIV-1 protease. Antiviral therapy; 2002: INT MEDICAL PRESS LTD 2-4 IDOL LANE, LONDON EC3R 5DD, ENGLAND.
138. Rhee S-Y, Taylor J, Fessel WJ, Kaufman D, Towner W, Troia P, et al. HIV-1 protease mutations and protease inhibitor cross-resistance. Antimicrobial agents and chemotherapy. 2010;54(10):4253-61.
139. Eastman PS, Mittler J, Kelso R, Gee C, Boyer E, Kolberg J, et al. Genotypic changes in human immunodeficiency virus type 1 associated with loss of suppression of plasma viral RNA levels in subjects treated with ritonavir (Norvir) monotherapy. Journal of virology. 1998;72(6):5154-64.
140. Nijhuis M, Schuurman R, de Jong D, Erickson J, Gustchina E, Albert J, et al. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. Aids. 1999;13(17):2349-59.
141. Mammano F, Trouplin V, Zennou V, Clavel F. Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug. Journal of virology. 2000;74(18):8524-31.
142. Clemente JC, Hemrajani R, Blum LE, Goodenow MM, Dunn BM. Secondary mutations M36I and A71V in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation D30N. Biochemistry. 2003;42(51):15029-35.
143. Molla A, Korneyeva M, Gao Q, Vasavanonda S, Schipper PJ, Mo H-M, et al. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. Nature medicine. 1996;2(7):760-6.
144. EBERLE J, BECHOWSKY B, ROSE D, HAUSER U, VON DER HELM K, GÜRTLER L, et al. Resistance of HIV type 1 to proteinase inhibitor Ro 31-8959. AIDS research and human retroviruses. 1995;11(6):671-6.
145. Jacobsen H, Yasargil K, Winslow DL, Craig JC, Kröhn A, Duncan IB, et al. Characterization of human immunodeficiency virus type 1 mutantswith decreased sensitivity to proteinase inhibitor Ro 31-8959. Virology. 1995;206(1):527-34.

146. Wu TD, Schiffer CA, Gonzales MJ, Taylor J, Kantor R, Chou S, et al. Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. Journal of virology. 2003;77(8):4836-47.
147. Doherty KM, Nakka P, King BM, Rhee S-Y, Holmes SP, Shafer RW, et al. A multifaceted analysis of HIV-1 protease multidrug resistance phenotypes. BMC bioinformatics. 2011;12(1):477.
148. Kempf DJ, Isaacson JD, King MS, Brun SC, Xu Y, Real K, et al. Identification of genotypic changes in human immunodeficiency virus protease that correlate with reduced susceptibility to the protease inhibitor lopinavir among viral isolates from protease inhibitor-experienced patients. Journal of virology. 2001;75(16):7462-9.
149. Condra JH, Holder DJ, Schleif WA, Blahy OM, Danovich RM, Gabryelski LJ, et al. Genetic correlates of in vivo viral resistance to indinavir, a human immunodeficiency virus type 1 protease inhibitor. Journal of virology. 1996;70(12):8270-6.
150. Condra JH, Schleif WA, Blahy OM, Gabryelski LJ, Graham DJ, Quintero J, et al. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. Nature. 1995;374(6522):569-71.
151. Jacobsen H, Hänggi M, Ott M, Duncan IB, Owen S, Andreoni M, et al. In vivo resistance to a human immunodeficiency virus type 1 proteinase inhibitor: mutations, kinetics, and frequencies. Journal of Infectious Diseases. 1996;173(6):1379-87.
152. Steegen K, Bronze M, Papathanasopoulos M, van Zyl G, Goedhals D, Van Vuuren C, et al. Prevalence of Antiretroviral Drug Resistance in Patients Who Are Not Responding to Protease Inhibitor–Based Treatment: Results From the First National Survey in South Africa. Journal of Infectious Diseases. 2016;214(12):1826-30.
153. Van Zyl GU, Van Mens TE, McIlleron H, Zeier M, Nachega JB, Decloedt E, et al. Low lopinavir plasma or hair concentrations explain second line protease inhibitor failures in a resource-limited setting. Journal of acquired immune deficiency syndromes (1999). 2011;56(4):333.
154. Zybarth G, Carter C. Domains upstream of the protease (PR) in human immunodeficiency virus type 1 Gag-Pol influence PR autoprocessing. Journal of virology. 1995;69(6):3878-84.
155. Fisher R, van Zyl GU, Travers SA, Pond SLK, Engelbrech S, Murrell B, et al. Deep sequencing reveals minor protease resistance mutations in patients failing a protease inhibitor regimen. Journal of virology. 2012;86(11):6231-7.
156. Levison JH, Orrell C, Gallien S, Kuritzkes DR, Fu N, Losina E, et al. Virologic failure of protease inhibitor-based second-line antiretroviral therapy without resistance in a large HIV treatment program in South Africa. PLoS ONE. 2012;7(3):e32144.
157. Berhanu R, Evans D, Ive P, Spencer D, Firnhaber C, Sanne I, et al., editors. Second line failure and protease inhibitor resistance in a clinic in Johannesburg, South Africa. Conference on Retroviruses and Opportunistic Infections Boston, USA; 2014.
158. El-Khatib Z, Ekström AM, Ledwaba J, Mohapi L, Laher F, Karstaedt A, et al. nViremia and drug resistance among HIV-1 patients on antiretroviral treatment–a cross-sectional study in Soweto, South Africa. AIDS (London, England). 2010;24(11):1679.
159. van Zyl GU, van der Merwe L, Claassen M, Zeier M, Preiser W. Antiretroviral resistance patterns and factors associated with resistance in adult patients failing NNRTI-based regimens in the western cape, South Africa. Journal of medical virology. 2011;83(10):1764-9.
160. Larrouy L, Chazallon C, Landman R, Capitant C, Peytavin G, Collin G, et al. Gag mutations can impact virological response to dual-boosted protease inhibitor combinations in antiretroviral-naive HIV-infected patients. Antimicrobial agents and chemotherapy. 2010;54(7):2910-9.
161. Zhang Y-M, Imamichi H, Imamichi T, Lane HC, Falloon J, Vasudevachari M, et al. Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. Journal of virology. 1997;71(9):6662-70.
162. Gatanaga H, Suzuki Y, Tsang H, Yoshimura K, Kavlick MF, Nagashima K, et al. Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. Journal of Biological Chemistry. 2002;277(8):5952-61.

163. Parry CM, Kohli A, Boinett CJ, Towers GJ, McCormick AL, Pillay D. Gag determinants of fitness and drug susceptibility in protease inhibitor-resistant human immunodeficiency virus type 1. Journal of virology. 2009;83(18):9094-101.
164. Dam E, Quercia R, Glass B, Descamps D, Launay O, Duval X, et al. Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. PLoS pathogens. 2009;5(3):e1000345.
165. Gupta RK, Kohli A, McCormick AL, Towers GJ, Pillay D, Parry CM. Full length HIV-1 gag determines protease inhibitor susceptibility within in vitro assays. AIDS (London, England). 2010;24(11):1651.
166. Fun A, Wensing AM, Verheyen J, Nijhuis M. Human Immunodeficiency Virus Gag and protease: partners in resistance. Retrovirology. 2012;9(1):63.
167. Doyon L, Croteau G, Thibeault D, Poulin F, Pilote L, Lamarre D. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. Journal of virology. 1996;70(6):3763-9.
168. Nijhuis M, Deeks S, Boucher C. Implications of antiretroviral resistance on viral fitness. Current opinion in infectious diseases. 2001;14(1):23-8.
169. Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, et al. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. Journal of clinical microbiology. 2005;43(1):406-13.
170. Mohamed S, Penaranda G, Gonzalez D, Camus C, Khiri H, Boulmé R, et al. Comparison of ultra-deep versus Sanger sequencing detection of minority mutations on the HIV-1 drug resistance interpretations after virological failure. AIDS. 2014;28(9):1315-24.
171. Tzou PL, Ariyaratne P, Varghese V, Lee C, Rakhmanaliev E, Villy C, et al. Comparison of an In Vitro Diagnostic Next-Generation Sequencing Assay with Sanger Sequencing for HIV-1 Genotypic Resistance Testing. Journal of clinical microbiology. 2018:JCM. 00105-18.
172. Ansorge WJ. Next-generation DNA sequencing techniques. New biotechnology. 2009;25(4):195-203.
173. Shendure J, Ji H. Next-generation DNA sequencing. Nature biotechnology. 2008;26(10):1135.
174. Illumina. NGS library prep methods2016 2018.
175. Illumina. Illumina Sequencing Technology 2016.
176. Alaeus A, Lidman K, Sönnerborg A, Albert J. Subtype-specific problems with quantification of plasma HIV-1 RNA. Aids. 1997;11(7):859-65.
177. Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, Schultze JL, et al. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. Scientific reports. 2018;8(1):10950.
178. Boltz VF, Zheng Y, Lockman S, Hong F, Halvas EK, McIntyre J, et al. Role of low-frequency HIV-1 variants in failure of nevirapine-containing antiviral therapy in women previously exposed to single-dose nevirapine. Proceedings of the National Academy of Sciences. 2011;108(22):9202-7.
179. Johnson JA, Li J-F, Wei X, Lipscomb J, Irlbeck D, Craig C, et al. Minority HIV-1 drug resistance mutations are present in antiretroviral treatment–naive populations and associate with reduced treatment efficacy. PLoS Med. 2008;5(7):e158.
180. Zain RM, Ibrahim N, Ismail S, Suppiah J, Rahim NAM, Thayan R. Identifications of drug resistance mutations among antiretroviral treatment naive HIV-1 patients in Peninsular Malaysia. Asian Pacific journal of tropical medicine. 2017;10(1):75-8.
181. Barber TJ, Harrison L, Asboe D, Williams I, Kirk S, Gilson R, et al. 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. 2012;67(4):995-1000.
182. Lecossier D, Shulman NS, Morand-Joubert L, Shafer RW, Joly V, Zolopa AR, et al. Detection of minority populations of HIV-1 expressing the K103N resistance mutation in patients failing nevirapine. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2005;38(1):37-42.

183. Li JZ, Kuritzkes DR. Clinical implications of HIV-1 minority variants. Clinical infectious diseases. 2013;56(11):1667-74.
184. Le T, Chiarella J, Simen BB, Hanczaruk B, Egholm M, Landry ML, et al. Low-abundance HIV drug-resistant viral variants in treatment-experienced persons correlate with historical antiretroviral use. PLoS ONE. 2009;4(6):e6079.
185. Lataillade M, Chiarella J, Yang R, DeGrosky M, Uy J, Seekins D, et al. Virologic failures on initial boosted-PI regimen infrequently possess low-level variants with major PI resistance mutations by ultra-deep sequencing. PLoS ONE. 2012;7(2):e30118.
186. McKinnon JE, Delgado R, Pulido F, Shao W, Arribas JR, Mellors JW. Single genome sequencing of HIV-1 gag and protease resistance mutations at virologic failure during the OK04 trial of simplified versus standard maintenance therapy. Antiviral therapy. 2010;16(5):725-32.
187. Dunn DT, Stöhr W, Arenas-Pinto A, Tostevin A, Mbisa JL, Paton NI. Next generation sequencing of HIV-1 protease in the PIVOT trial of protease inhibitor monotherapy. Journal of Clinical Virology. 2018;101:63-5.
188. Dykes C, Najjar J, Bosch RJ, Wantman M, Furtado M, Hart S, et al. Detection of drug-resistant minority variants of HIV-1 during virologic failure of indinavir, lamivudine, and zidovudine. The Journal of infectious diseases. 2004;189(6):1091-6.
189. Perrier M, Visseaux B, Landman R, Joly V, Todesco E, Yazdanpanah Y, et al. No impact of HIV-1 protease minority resistant variants on the virological response to a first-line PI-based regimen containing darunavir or atazanavir. Journal of Antimicrobial Chemotherapy. 2017;73(1):173-6.
190. Zhou Z, Wagar N, DeVos JR, Rottinghaus E, Diallo K, Nguyen DB, et al. Optimization of a low cost and broadly sensitive genotyping assay for HIV-1 drug resistance surveillance and monitoring in resource-limited settings. PLoS ONE. 2011;6(11):e28184.
191. DeAngelis MM, Wang DG, Hawkins TL. Solid-phase reversible immobilization for the isolation of PCR products. Nucleic acids research. 1995;23(22):4742.
192. Borgström E, Lundin S, Lundeberg J. Large scale library generation for high throughput sequencing. PLoS ONE. 2011;6(4):e19119.
193. Larkin MA, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. bioinformatics. 2007;23(21):2947-8.
194. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in bioinformatics. 2008;9(4):299-306.
195. Prism G. GraphPad Prism. version 7. La Jolla, CA: GraphPad Software. 2008.
196. Wallis CL, Mellors JW, Venter WD, Sanne I, Stevens W. Protease inhibitor resistance is uncommon in HIV-1 subtype C infected patients on failing second-line lopinavir/r-containing antiretroviral therapy in South Africa. AIDS research and treatment. 2010;2011.
197. Walmsley S, Bernstein B, King M, Arribas J, Beall G, Ruane P, et al. Lopinavir–ritonavir versus nelfinavir for the initial treatment of HIV infection. New England Journal of Medicine. 2002;346(26):2039-46.
198. Todd J, Pachl C, White R, Yeghiazarian T, Johnson P, Taylor B, et al. Performance characteristics for the quantitation of plasma HIV-1 RNA using branched DNA signal amplification technology. Journal of acquired immune deficiency syndromes and human retrovirology: official publication of the International Retrovirology Association. 1995;10:S35-44.
199. Yu K, Xing J, Zhang J, Zhao R, Zhang Y, Zhao L. Effect of multiple cycles of freeze–thawing on the RNA quality of lung cancer tissues. Cell and tissue banking. 2017;18(3):433-40.
200. Panaccio M, Lew A. PCR based diagnosis in the presence of 8%(v/v) blood. Nucleic acids research. 1991;19(5):1151.
201. Klein A, Barsuk R, Dagan S, Nusbaum O, Shouval D, Galun E. Comparison of methods for extraction of nucleic acid from hemolytic serum for PCR amplification of hepatitis B virus DNA sequences. Journal of clinical microbiology. 1997;35(7):1897-9.

202. Levere RD, Gong Y-F, Kappas A, Bucher DJ, Wormser GP, Abraham NG. Heme inhibits human immunodeficiency virus 1 replication in cell cultures and enhances the antiviral effect of zidovudine. Proceedings of the National Academy of Sciences. 1991;88(5):1756-9.
203. Varghese V, Wang E, Babrzadeh F, Bachmann MH, Shahriar R, Liu T, et al. Nucleic acid template and the risk of a PCR-Induced HIV-1 drug resistance mutation. PLoS ONE. 2010;5(6):e10992.
204. Gianella S, Delport W, Pacold ME, Young JA, Choi JY, Little SJ, et al. Detection of minority resistance during early HIV-1 infection: natural variation and spurious detection rather than transmission and evolution of multiple viral variants. Journal of virology. 2011:JVI. 02582-10.
205. Dudley DM, Bailey AL, Mehta SH, Hughes AL, Kirk GD, Westergaard RP, et al. Cross-clade simultaneous HIV drug resistance genotyping for reverse transcriptase, protease, and integrase inhibitor mutations by Illumina MiSeq. Retrovirology. 2014;11(1):122.
206. Allhoff M, Schönhuth A, Martin M, Costa IG, Rahmann S, Marschall T, editors. Discovering motifs that induce sequencing errors. BMC bioinformatics; 2013: BioMed Central.
207. Shao W, Boltz VF, Spindler JE, Kearney MF, Maldarelli F, Mellors JW, et al. Analysis of 454 sequencing error rate, error sources, and artifact recombination for detection of Low-frequency drug resistance mutations in HIV-1 DNA. Retrovirology. 2013;10(1):18.
208. Mild M, Hedskog C, Jernberg J, Albert J. Performance of ultra-deep pyrosequencing in analysis of HIV-1 pol gene variation. PLoS ONE. 2011;6(7):e22741.
209. Vandenbroucke I, Van Marck H, Verhasselt P, Thys K, Mostmans W, Dumont S, et al. Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications. Biotechniques. 2011;51(3):67-77.
210. Poon AF, Swenson LC, Dong WW, Deng W, Kosakovsky Pond SL, Brumme ZL, et al. Phylogenetic analysis of population-based and deep sequencing data to identify coevolving sites in the nef gene of HIV-1. Molecular biology and evolution. 2009;27(4):819-32.
211. Liu S-L, Rodrigo AG, Shankarappa R, Learn GH, Hsu L, Davidov O, et al. HIV quasispecies and resampling. Science. 1996;273(5274):413-7.
212. Boltz VF, Coomer C, Mellors J, Kearney M, Coffin J, editors. Analysis of resistance haplotypes using primer IDs and next gen sequencing of HIV RNA. Conference on retroviruses and opportunistic infections; 2015.
213. Nijhuis M, Wensing AM, Bierman WF, de Jong D, Kagan R, Fun A, et al. Failure of treatment with first-line lopinavir boosted with ritonavir can be explained by novel resistance pathways with protease mutation 76V. The Journal of infectious diseases. 2009;200(5):698-709.
214. Young TP, Parkin NT, Stawiski E, Pilot-Matias T, Trinh R, Kempf DJ, et al. Prevalence, mutation patterns, and effects on protease inhibitor susceptibility of the L76V mutation in HIV-1 protease. Antimicrobial agents and chemotherapy. 2010;54(11):4903-6.
215. Mueller S, Daeumer M, Kaiser R, Walter H, Colonno R, Korn K, editors. Susceptibility to saquinavir and atazanavir in highly protease inhibitor (PI) resistant HIV-1 is caused by lopinavir-induced drug resistance mutation L76V. Antiviral therapy; 2004: INT MEDICAL PRESS LTD 2-4 IDOL LANE, LONDON EC3R 5DD, ENGLAND.
216. Croteau G, Doyon L, Thibeault D, McKercher G, Pilote L, Lamarre D. Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors. Journal of virology. 1997;71(2):1089-96.
217. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. Journal of molecular biology. 1994;235(3):983-1002.
218. Kagan RM, Shenderovich MD, Heseltine PN, Ramnarayan K. Structural analysis of an HIV-1 protease I47A mutant resistant to the protease inhibitor lopinavir. Protein Science. 2005;14(7):1870-8.
219. Watkins T, Resch W, Irlbeck D, Swanstrom R. Selection of high-level resistance to human immunodeficiency virus type 1 protease inhibitors. Antimicrobial agents and chemotherapy. 2003;47(2):759-69.

220. Collins JA, Thompson MG, Paintsil E, Ricketts M, Gedzior J, Alexander L. Competitive fitness of nevirapine-resistant human immunodeficiency virus type 1 mutants. Journal of virology. 2004;78(2):603-11.
221. Polilli E, Parruti G, Cosentino L, Sozio F, Saracino A, Consorte A, et al. Rapid and persistent selection of the K103N mutation as a majority quasispecies in a HIV1-patient exposed to efavirenz for three weeks: a case report and review of the literature. Journal of medical case reports. 2009;3(1):9132.
222. Walsh JC, Pozniak AL, Nelson MR, Mandalia S, Gazzard BG. Virologic rebound on HAART in the context of low treatment adherence is associated with a low prevalence of antiretroviral drug resistance. Journal of acquired immune deficiency syndromes (1999). 2002;30(3):278-87.
223. Bangsberg DR, Hecht FM, Charlebois ED, Zolopa AR, Holodniy M, Sheiner L, et al. Adherence to protease inhibitors, HIV-1 viral load, and development of drug resistance in an indigent population. Aids. 2000;14(4):357-66.
224. Myint L, Matsuda M, Matsuda Z, Yokomaku Y, Chiba T, Okano A, et al. Gag non-cleavage site mutations contribute to full recovery of viral fitness in protease inhibitor-resistant human immunodeficiency virus type 1. Antimicrobial agents and chemotherapy. 2004;48(2):444-52.
225. Verheyen J, Litau E, Sing T, Daumer M, Balduin M, Oette M, et al. Compensatory mutations at the HIV cleavage sites p7/p1 and p1/p6-gag in therapy-naive and therapy-experienced patients. Antiviral therapy. 2006;11(7):879.
226. Kaye JF, Lever AM. Nonreciprocal packaging of human immunodeficiency virus type 1 and type 2 RNA: a possible role for the p2 domain of Gag in RNA encapsidation. Journal of virology. 1998;72(7):5877-85.
227. Lambert-Niclot S, Flandre P, Malet I, Canestri A, Soulié C, Tubiana R, et al. Impact of gag mutations on selection of darunavir resistance mutations in HIV-1 protease. Journal of Antimicrobial Chemotherapy. 2008;62(5):905-8.

6.Appendices

Appendix 1: Ethical clearance

R14/49 «Tit init name»

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M1704127

NAME: (Principal Investigator)	Ms Mhlekazi Molatoli and Dr Sergio Carmona
DEPARTMENT:	Molecular Medicine and Haematology National Health Laboratory Services Charlotte Maxeke Johannesburg Academic Hospital
PROJECT TITLE:	Detection of Antiretroviral Drug Resistant Minority Variants in Patients Falling Second-Line Antiretroviral Drug Therapy in South Africa
DATE CONSIDERED:	Adhoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-Study (M120254)
SUPERVISOR:	Prof Maria Papathanasopoulos and Dr Kim Steegen
APPROVED BY:	Professor P. Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	12/05/2017

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

Appendix 2:The demographic and clinical data of all 188 participants included in this study.

Participant ID	VL	VL log_{10}	Current CD4	Current ART	Previous ART	Period (Months)	Age (Years)	Gender
ECC005	1322	3.1	139	3TC-TDF-LPV/r	d4T-3TC-NVP	82	45	F
ECC006	20745	4.3	325	AZT-3TC-LPV/r	d4T-3TC-EFV	47	39	F
ECC009	95461	5.0	349	AZT-3TC-LPV/r	TDF-3TC-EFV	39	39	M
ECC010	362141	5.6	121	3TC-TDF-LPV/r	AZT-ddI-LPV/r	10	38	F
ECC011	22320	4.3	369	TDF-3TC-LPV/r	d4T-3TC-EFV	31	38	F
ECC013	38501	4.6	263	3TC-TDF-LPV/r	d4T-3TC-NVP	105	43	F
ECC014	4612	3.7	482	TDF-3TC-LPV/r	d4T-3TC-EFV	12	29	F
ECC017	227848	5.4	209	AZT-3TC-LPV/r	3TC-d4T-EFV	20	36	M
ECC018	1725878	6.2	37	3TC-TDF-LPV/r	AZT-ddI-d4T	7	37	F

ECC021	18050	4.3	88	TDF-3TC-LPV/r	D4T-3TC-EFV	37	35	M
ECC029	171703	5.2	73	3TC-TDF-LPV/r	d4T-3TC-EFV	11	40	M
ECC030	165486	5.2	298	TDF-3TC-LPV/r	TDF-AZT-3TC-EFV	29	32	F
ECC031	48759	4.7	984	3TC-ABC-LPV/r	d4T-3TC-EFV	32	44	F
GPC001	1669	3.2	NA	AZT-3TC-LPV/r	NA	NA	37	M
GPC002	406339	5.6	191	EFV-LPV/r	d4T-3TC-EFV	34	31	F
GPC004	2631095	6.4	86	AZT-3TC-TDF-ATZ/r	AZT-3TC-TDF-LPV/r	21	41	M
GPC005	2721	3.4	119	AZT-3TC-LPV/r	D4T-3TC-EFV	35	42	M
GPC011	143360	5.2	261	TDF-3TC-LPV/r	AZT-3TC-EFV	NA	54	F
GPC014	26484	4.4	209	d4T-3TC-LPV/r	d4T-3TC-TDF-EFV	15	33	F
GPC015	470866	5.7	203	3TC-TDF-LPV/r	NA	65	41	F

GPC016	3985	3.6	333	TDF-3TC-LPV/r	d4T-3TC EFV-LPV/r-ATZ/r	26	54	F
GPC017	2190	3.3	568	EFV-3TC-LPV/r	D4T-3TC-EFV	30	40	F
GPC019	96363	5.0	382	TDF-3TC-LPV/r	AZT-DDI-D4T-EFV-LPV/r	32	47	M
GPC026	116567	5.1	211	d4T-3TC-LPV/r	AZT-3TC-LPV/r d4T-3TC-EFV	9	29	F
GPC030	879270	5.9	34	AZT-3TC-TDF- LPV/r	d4T-3TC-EFV	7	39	F
GPC032	8426	3.9	56	AZT-ddI-3TC-LPV/r	d4T-3TC-EFV-NVP	11	19	F
GP038	38849	4.6	264	3TC-TDF-LPV/r	d4T-3TC-EFV 3TC-TDF-LPV/r	16	42	F
GPC041	239660	5.4	192	AZT-3TC-LPV/r	ddI-TDF-LPV	23	38	F
GPC043	391441	5.6	319	TDF-3TC-LPV/r	TDF-3TC-NVP	13	31	F
GPC046	242545	5.4	215	AZT-3TC-LPV/r	TDF-3TC-EFV	17	33	F

GPC048	18422	4.3	254	AZT-3TC-LPV/r	TDF-FTC-EFV	9	19	F
GPC049	10325	4.0	160	AZT-3TC-TDF-LPV/r	3TC-TDF-EFV-d4T-AZT	49	39	F
GPC052	172637	5.2	119	TDF-3TC-LPV/r	d4T-3TC-EFV	34	52	M
GPC055	28996	4.5	151	AZT-3TC-LPV/r	TDF-3TC-NVP	9	26	F
GPC056	13762	6.1	53	TDF-3TC-LPV/r	D4T-3TC-EFV	30	43	F
GPC057	59751	4.8	332	d4T-3TC-LPV/r	d4T-3TC-EFV	17	33	M
GPC058	97975	5	159	AZT-ddI-LPV/r	d4T-3TC-NVP	7	36	F
GPC061	232154	4.4	6	AZT-3TC-LPV/r	d4T-3TC-EFV	15	27	F
GPC071	53000	4.7	68	TDF-FTC-ATZ/R	DDI-EFV-	24	43	F
GPC076	115528	5.1	149	TDF-3TC-LPV/r	NA	30	33	F
GPC080	26007	4.4	315	TDF-3TC-LPV/r	d4T-3TC-NVP	29	36	F

GPC090	82839	4.9	127	AZT-3TC-LPV/r	3TC-TDF-EFV	7	46	M
GPC093	10370	4.0	541	3TC-TDF-LPV/r	NA	60	62	F
GPC094	3477351	6.5	167	3TC-TDF-LPV/r	AZT-3TC-LPV/r	11	56	M
GPC117	179352	5.3	190	TDF-LPV/r-3TC	D4T-3TC-EFV-TDF-LPV/r	22	31	F
KZC005	7452	3.9	180	3TC-TDF-LPV/r	d4T-3TC-NVP	26	27	F
KZC007	NA	NA	NA	AZT-ddI-LPV/r	d4T-3TC-EFV	51	38	F
KZC008	NA	NA	NA	3TC-LPV/r	3TC-TDF-EFV	12	35	F
KZC009	8161	3.9	269	TDF-3TC-LPV/r	d4T-3TC-NVP	41	40	F
KZC010	5179	3.7	383	3TC+ABC+LPV/r	AZT+3TC+EFV	12	31	F
KZC012	25729	4.4	41	ABC-3TC-LPV/r	d4T-3TC-EFV	NA	18	F
KZC013	11195	4.0	7	AZT-3TC-LPV/r	3TC-TDF-NVP	9	34	F

KZC018	285247	5.5	458	TDF-3TC-LPV/r	NA	28	59	F
KZC019	88130	4.9	183	TDF-3TC-LPV/r	d4T-3TC-EFV	32	38	F
KZC021	26691	4.4	59	AZT-3TC-LPV/r	3TC-TDF-NVP	21	38	M
KZC023	4204	3.6	356	3TC-TDF-LPV/r	NA	61	25	F
KZC024	109352	5.0	30	3TC-TDF-LPV/r	3TC-TDF-d4T-NVP	NA	32	F
KZC025	53124	4.7	167	AZT-3TC-LPV/r	AZT-ddI-NVP	41	39	F
KZC026	467608	5.7	220	AZT-3YC-PV/r	d4T-3TC-EFV	15	49	M
KZC028	328544	5.5	48	AZT-3TC-LPV/r	AZT-ddI-LPV/r d4T-3TC-NVP	33	31	F
KZC032	622500	5.8	49	TDF-3TC-LPV/r	d4T-3TC-EFV	10	34	F
KZC034	52563	4.7	222	AZT-ddI-LPV/r	ABC-d4T-NVP	26	44	M
KZC035	243026	5.4	18	AZT-ddI-LPV/r	d4T-ABC-NVP	26	44	F

KZC038	189466	5.3	49	AZT-TDF-3TC-LPV/r	NA	43	35	F
KZC040	118455	5.1	36	AZT-3TC-LPV/r	d4T-3TC-NVP	17	39	F
KZC042	51111	4.7	358	3TC-TDF-LPV/r	d4T-3TC-EFV	7	47	M
KZC045	5000	3.7	576	AZT+ddI+LPV/r	d4T+3TC+NVP	24	40	F
KZC046	89559	5.0	NA	TDF-3TC-LPV/r	NA	NA	31	F
KZC044	393690	5.6	93	ddI-ABC-LPV/r	d4T-3TC-EFV	43	18	F
KZC045	5000	3.7	576	AZT-ddI-LPV/r	d4T-3TC-NVP	24	40	F
KZC055	8980	4.0	239	TDF-3TC-LPV/r	d4T-3TC-EFV d4T-3TC-NVP	11	34	F
KZC054	96642	5.0	138	3TC-TDF-LPV/r	AZT-3TC-EFV AZT-3TC-NVP	25	29	F
KZC057	7336	3.9	265	TDF-3TC-LPV/r	D4T-3TC-NPV	NA	35	F
KZC059	87346	4.9	148	ABC-3TC-LPV/r	TDF-3TC-EFV	25	42	F

KZC060	371228	5.6	204	TDF-3TC-LPV/r	d4T-3TC-NPV-EFV	38	60	M
KZC061	121983	5.1	285	TDF-3TC-LPV/r	d4T-3TC-EFV	8	35	M
KZC063	12433	4.1	270	TDF-3TC-LPV/r	d4T-3TC-EFV	8	34	F
KZC064	205670	5.3	120	TDF-3TC-LPV/r	d4T-3TC-EFV	8	39	F
KZC066	18428	4.3	288	AZT-ddI-LPV/r	d4T-3TC-EFV	45	18	M
KZC067	5883	3.8	293	TDF-3TC-LPV/r	NA	17	50	M
KZC069	14542	4.2	313	3TC-TDF-LPV/r	d4T-3TC-NVP	24	42	F
KZC070	952662	6.0	86	TDF-3TC-LPV/r	NA	17	44	M
KZC071	4629	3.7	24	TDF-3TC-LPV/r	D4T-3TC-NVP	34	32	M
KZC072	423858	5.6	49	TDF-3TC-LPV/r	d4T-3TC-EFV	10	42	F
KZC073	1629	3.2	322	TDF-AZT-LPV/r	DDI-3TC-EFV	58	37	F

KZC074	73664	4.9	231	3TC-TDF-LPV/r	d4T-3TC-EFV	39	33	M
KZC075	NA	NA	393	AZT-3TC-LPV/r	TDF-3TC-EFV	20	53	M
KZC076	5550	3.7	345	AZT-3TC-LPV/r	d4T-3TC-NVP	74	38	F
KZC078	35931	4.6	254	AZT-ddI-LPV/r	d4T-3TC-NVP	9	45	M
KZC080	314120	5.5	337	AZT-ddI-LPV/r	d4T-3TC-EFV	76	64	F
KZC084	122608	5.1	152	ABC-3TC-LPV/r	d4T-3TC-EFV	58	20	F
KZC089	61181	4.3	352	TDF-3TC-LPV/r	TDF-3TC-NVP	23	38	F
KZC092	251478	5.4	10	3TC-TDF-LPV/r	AZT-3TC-EFV	22	52	M
KZC093	13895	4.1	303	3TC+TDF+LPV/r	d4T+3TC+EFV	9	47	F
KZC094	269493	5.4	67	3TC-TDF-LPV/r	d4T-3TC-NVP	10	25	F
KZC093	10308	4.0	303	TDF-3TC-LPV/r	d4T-3TC-EFV	32	47	F

KZC095	85779	4.9	77	TDF-3TC-LPV/r	NA	32	51	F
KZC096	38318	4.4	141	TDF-3TC-LPV/r	D4T-3TC-EFV	12	52	M
KZC098	112512	5.1	171	TDF-3TC-LPV/r	TDF-3TC-EFV	24	44	M
KZC100	12408	4.1	487	TDF-3TC-LPV/r	d4T-3TC-EFV	NA	40	f
KZC101	8518	3.9	700	ddI-3TC-LVP/r	d4T-3TC-NVP	71	41	F
KZC102	31333	4.5	579	3TC-TDF-LPV/r	d4T-3TC-EFV	14	41	F
KZC103	95885	5.0	38	TDF-3TC-LPV/r	d4T-3TC-NVP	72	25	F
KZC108	8106	3.9	318	TDF-3TC-LPV/r	d4T-3TC-EFV	7	46	F
KZC109	18407	4.3	38	AZT-3TC-LPV/r	d4T-3TC-NVP	35	45	F
KZC110	5524	3.7	316	3TC-TDF-LPV/r	d4T-3TC-NVP	10	40	F
KZC112	57871	4.8	286	TDF-3TC-LPV/r	AZT-3TC-EFV	14	49	M

KZC113	388677	5.6	107	AZT-3TC-LPV/r	TDF-3TC-NVP	9	28	F
KZC116	54710	4.7	139	3TC-TDF-LPV/r	d4T-3TC-EFV	10	49	M
KZC117	5157	3.7	261	TDF-3TC-LPV/r	S4T-3TC-EFV	10	48	F
KZC118	58722	4.8	92	TDF-3TC-LPV/r	d4T-3TC-NVP	14	20	F
KZC120	20390	4.3	208	AZT-3TC-LPV/r	d4T-3TC-EFV	15	54	M
KZC121	34558	4.5	743	AZT-3TC-LPV/r	d4T-3TC-NVP	NA	39	F
KZC124	9456	4.0	589	TDF-3TC-LPV/r	d4T-3TC-NVP	22	29	F
KZC126	100473	5.0	63	3TC-TDF-LPV/r	d4T-3TC-NVP	22	34	F
KZC127	941846	6	53	AZT-3TC-LPV/r	TDF-3TC-EFV	9	40	M
KZC130	14700	4.2	129	TDF-3TC-LPV/r	d4T-3TC-EFV	7	46	F
KZC132	72071	4.9	211	AZT-3TC-LPV/r	NA	39	32	F

KZC133	104661	5.0	171	AZT-3TC-LPV/r	d4T-3TC-EFV	31	39	F
KZC134	2574217	6.4	110	ABC-3TC-LPV/r	AZT-3TC-LPV/r d4T-3TC-EFV	3	33	F
KZC135	6922	3.8	105	AZT-3TC-LPV/r	TDF-3TC-EFV	20	55	F
KZC137	199820	5.3	161	AZT-3TC-LPV/r	d4T-3TC-NVP	46	35	M
KZC138	1715	3.2	721	TDF-3TC-LPV/r	d4T-3TC-SQV/r	6	23	F
KZC139	9923	4.0	433	3TC-TDF-LPV/r	d4T-3TC-EFV	30	42	F
KZC140	35408	4.5	175	TDF-3TC-LPV/r	d4T-3TC-NVP	16	50	M
KZC143	3656	3.6	189	AZT-3TC-LPV/r	d4T-3TC-NVP	96	40	F
KZC146	15224	4.2	286	FTC-LPV/r	D4T-3TC-EFV	11	20	M
KZC147	NA	NA	253	AZT-3TC-LPV/r	d4T+3TC+EFV	36	55	F
KZC148	138662	5.1	318	TDF-3TC-LPV/r	d4T-3TC-EFV	11	41	F

KZC149	203637	5.3	6	TDF-3TC-LPV/r	D4T-TC-EFV	14	31	F
KZC151	11286	4.1	273	TDF-3TC-LPV/r	TDF-FTC-NVP	25	30	F
KZC154	7023	3.8	211	TDF-3TC-LPV/r	d4T-3TC-EFV	12	37	F
KZC155	922	3	275	3TC-TDF-LPV/r	d4T-3TC-EFV	90	50	M
KZC156	23528	4.4	367	3TC-TDF-LPV/r	d4T-3TC-NVP	12	30	F
KZC157	475399	5.9	48	3TC-AZT-LPV/r	3TC-TDF-NVP	9	28	F
KZC158	79971	4.9	300	3TC-TDF-LPV/r	3TC-TDF-EFV	27	40	F
KZC160	101164	5.0	71	3TC-TDF-LPV/r	d4T-3TC-NVP	42	37	F
KZC162	24159	4.4	137	TDF-3TC-LPV/r	ABC-3TC-EFV-d4T	NA	19	M
KZC168	83243	4.9	87	d4T-3TC-LPV/r	d4T-3TC-EFV	22	48	M
KZC172	28750	4.5	531	AZT-3TC-LPV/r	3TC-TDF-EFV	7	31	F

KZC174	599208	5.8	45	ABC-3TC-LPV/r	D4T-3TC-EFV	13	39	F
KZC181	22735	4.4	231	TDF-3TC-LPV/r	d4T-3TC-EFV	24	18	F
KZC182	16107	4.2	235	3TC-TDF-LPV/r	d4T-3TC-EFV	NA	33	F
KZC183	99092	5.0	NA	TDF-3TC-LPV/r	d4T-3TC-EFV	13	33	F
KZC189	2325	3.4	295	TDF-3TC-LPV/r	d4T-3TC-NVP	NA	32	F
KZC190	2116	3.3	287	3TC-LPV/r	NA	40	26	F
KZC192	117133	5.1	NA	TDF-3TC-LPV/r	d4T-3TC-EFV	29	48	F
KZC193	4791	3.7	391	TDF-3TC-LPV/r	d4T-3TC-EFV	31	45	M
KZC195	2157	3.3	214	AZT-3TC-LPV/r	TDF-3TC-NVP	8	30	F
KZC200	43195	4.6	480	AZT-3TC-LPV/r	d4T-3TC-EFV	17	40	F
KZC201	4331	3.6	120	TDF-3TC-LPV/r	d4T-3TC-EFV	64	20	M

KZC202	4740	3.7	491	AZT-FTC-LPV/r	d4T-3TC-NVP	14	18	F
KZC215	6635	3.8	103	AZT-3TC-LPV/r	d4T-3TC-NVP	55	30	F
MPC006	36355	4.6	142	AZT-3TC-LPV/r	d4T-3TC-EFV	17	47	M
MPC021	263645	5.4	205	AZT-3TC-LPV/r	TDF-3TC-EFV	18	47	M
MPC024	1555543	6.2	133	AZT-3TC-LPV/r	d4T-3TC-EFV	13	57	F
MPC025	69256	4.8	NA	AZT-3TC-LPV/r	d4T-AZT-TDF-3TC-EFV	23	34	F
MPC031	62203	4.8	79	TDF-3TC-LVP/r	NA	35	24	F
NWC003	4398	3.6	169	AZT-3TC-LPV/r	d4T-3TC-EFV	61	32	F
NWC004	167692	5.2	250	3TC-TDF-LPV/r	d4T-3TC-EFV	29	43	F
NWC005	37995	4.6	138	AZT-3TC-LPV/r	TDF-3TC-d4T-EFV	18	57	F
NWC007	439936	5.6	209	3TC-TDF-LPV/r	d4T-3TC-EFV	18	38	F

NWC008	515176	5.7	189	3TC-TDF-LPV/r	d4T-3TC-EFV	NA	38	F
NWC009	221535	5.3	36	3TC-TDF-LPV/r	NA	9	38	M
NWC010	1158556	6.1	45	TDF-3TC-LPV/r	D4T-3TC-EFV	30	27	F
NWC011	1429099	6.2	149	AZT-3TC-LPV/r	3TC-TDF-EFV	NA	50	F
NWC013	154302	5.2	194	AZT-3TC-LPV/r	TDF-3TC-EFV	14	39	F
NWC014	18923	4.3	8	ABC-3TC-LPV/r	TDF-3TC-d4T-EFV	20	32	M
NWC016	37378	4.6	159	TDF-3TC-LPV/r	d4T-3TC-EFV	19	65	M
NWC018	101468	5.0	183	AZT-3TC-LPV/r	d4T-3TC-EFV	23	33	F
NWC019	1397802	6.1	227	AZT-3TC-LPV/r	d4T-3TC-EFV	61	42	F
NWC020	29365	4.5	NA	AZT-3TC-LPV/r	d4T-3TC-EFV	61	51	F
NWC026	711688	5.9	5	TDF-3TC-LPV/r	d4T-3TC-EFV	NA	45	F

NWC028	78968	4.9	332	AZT-3TC-LPV/r	NA	88	51	M
NWC031	5042	3.7	228	AZT-ddI-LPV/r	d4T-3TC-EFV	70	53	F
NWC032	275689	5.4	541	AZT-3TC-LPV/r	3TC-TDF-EFV d4T-3TC-EFV	18	41	F
NWC033	315244	5.5	220	3TC-TDF-LPV/r	3TC-TDF-NVP	12	26	F
NWC034	96436	5.0	NA	3TC+TDF+LPV/r	AZT+ddI+d4T+EFV	73	62	M
NWC036	865701	5.9	142	AZT-3TC-LPV/r	TDF-3TC-EFV	43	24	M
NWC037	99748	5.0	73	AZT-3TC-LPV/r	d4T-3TC-EFV	61	36	F
NWC038	1692030	6.2	212	TDF-3TC-LPV/r	D4T-3TC-EFV	11	54	F
NWC040	16107	4.2	NA	TDF-3TC-LPV/r	d4T-3TC-EFV	NA	26	F
NWC041	767905	5.9	115	AZT-3TC-LPV/r	3TC-TDF-EFV	30	36	F
NWC042	1751490	6.2	62	TDF-3TC-LPV/r	D4T-3TC-EFV	53	26	F

LPC004	301446	5.5	NA	AZT-ddI-LPV/r	AZT-3TC-NVP	14	29	F
LPC005	20749	4.3	NA	TDF-3TC-LPVr	NA	44	38	F
LPC006	5784	3.8	NA	TDF-3TC-LPVr	d4T-3TC-EFC	46	38	F
LPC008	27156	4.4	205	AZT-3TC-LPV/r	3TC-TDF-EFV d4T-3TC-NVP	13	26	F

Appendix 3:Genotypic resistance profiles for 17 of the 46 participants with HIV-1 harbouring minority RT inhibitor mutations, for which there was no resultant change in the resistance levels to any of the ARV drugs.Highlighted in bold are participant sample sequences with atleast one mutations detected by only one analysis software

Participant ID	Sequencing Method	RT mutations		ARV drugs				
		Deepchek® (%); Geneious® (%)		ABC	TDF	FTC	AZT	EFV
		NRTI	NNRTI					
ECC014	Sanger	ND*	K103N	S	S	S	S	HLR
	NGS	ND*	K103N (83.0;88.0) E138Q (13.4; 3.7)	S	S	S	S	HLR
ECC030	Sanger	ND*	K103N	S	S	S	S	HLR
	NGS	ND*	K103N (26.5;15.3), Y181C (9.4;7.2), G190A (8.54;9.6)	S	S	S	S	HLR
GPC048	Sanger	ND*	K103N	S	S	S	S	HLR
	NGS	ND*	K103S (31.0;33.6), K103N (20.0;20.0), K103R (23.0;23.2), V106M (14.0;15.4)	S	S	S	S	HLR
GPC080	Sanger	K70E, M184V	V106M, V179D, Y181C, G190A, F227L	IR	LLR	HLR	S	HLR
	NGS	D67G (11.4;11.6), K70E (97.9;99.2), M184V (98.6;99.6)	V106M (97.6;99.3), V179D (74.1;79.7), Y181C (6.2;4.2) G190A (98.8;99.6), F227L (77.2;0.0)	IR	LLR	HLR	S	HLR
KZC026	Sanger	M184V	A98G, K103N, V179D, E138K	LLR	S	HLR	S	HLR

	NGS	M184V (98.0;99.9)	A98G (66.0;72.9), K103N (27.0;18.7), V106M (6.0;4.7), E138K (19.0;19.3), V179D (36.0;42.0)	LLR	S	HLR	S	HLR
KZC028	Sanger	M184V	A98G, K103N	LLR	S	HLR	S	HLR
	NGS	M184V (79.0;82.6)	A98G (92.0;96.8), K103N (61.0;64.6), E138G (6.0;0.0)	LLR	S	HLR	S	HLR
KZC034	Sanger	D67N, K70R, M184V, K219E,	K103N	HLR	LLR	HLR	IR	HLR
	NGS	D67N (98.9;99.9), K70R (99.5;999), M184V (99.4;99.9), K219E (99.6;99.9)	K103N (83.6;84.5) K103S (15.8;15.7)	HLR	LLR	HLR	IR	HLR
KZC121	Sanger	M41L, D67N, M184V, T215Y, K219E	ND*	HLR	IR	HLR	HLR	S
	NGS	M41L (99.3 ; 99.9), D67N (99.1;99.9), M184V (98.1 ;99.8), T215Y (99.4;99.8), K219E (94.8;0.0), K219Q (4.6;3.8)	ND*	HLR	IR	HLR	HLR	S
NWC003	Sanger	M184V	K101P, K103NS, E138K	LLR	S	HLR	S	HLR
	NGS	M184V (98.0;100.0)	K101P (98.0;99.9), K103N (6.0;6.0) K103S (90.0;94.4), E138K (98.0;99.9)	LLR	S	HLR	S	HLR
NWC018	Sanger	M184V	K103N, Y181C K238T	LLR	S	HLR	S	HLR
	NGS	M184V (98.0;99.9)	K103N (98.0;99.9), Y181C (98.0;99,9), A98G (10.0;8.2), P225H (5.0;5.2) K238T (21.0;0.0)	LLR	S	HLR	S	HLR
KZC147	Sanger	M41L, D67N, K70R, M184V, T215F, K219E	K103N, V108I	IR	IR	PLLR	HLR	HLR

	NGS	M41L (99.0;99.5), D67N (99.0;99.7), K70R (85.2;93.8), M184V (99.2;99.9), T215F (98.9;.0.0), K219E (99.5;0.0)	K103N (98.4;99.8) V108I (17.8;13.5)	IR	IR	PLLR	HLR	HLR
KZC162	Sanger	M184V	A98A, K103N, P225H	LLR	S	HLR	S	HLR
	NGS	M184V (83;90.4)	A98G (76.0;80.9), K103N (73.0;76.3), P225H (83.0;87.9)	LLR	S	HLR	S	HLR
KZC200	Sanger	M184V	K103S, G190A N348IN	LLR	S	HLR	S	S
	NGS	M184V (99.4 ; 99.9)	K103S (87.0 ; 88.3), E138K (22.0 ; 18.8) E138Q (10.7 ; 11.9), G190A (85.6 ; 85.6) N348IN (91.0 ; 0.0)	S	S	S	S	HLR
LPC005	Sanger	D67N, K70E, M184V	ND*	S	S	S	S	HLR
	NGS	D67N (91.1;91.4), D67G (8.3;8.3) K70E (99.3;99.3), M184V (99.2;99.9)	ND*	S	S	S	S	HLR
KZC059	Sanger	ND*	L100I, K103N	S	S	S	LLR	HLR
	NGS	T215A (9.5;7.8)	L100I (98.5;99.3), K103N (98.6;99.4)	S	S	S	LLR	HLR
NWC019	Sanger	M184V	Y188L	LLR	S	HLR	S	HLR
	NGS	M184V (98.0;99.8)	Y188L (96.0;99.2)	LLR	S	HLR	S	HLR

NWC028	Sanger	ND*	K103N	S	S	S	S	HLR
	NGS	ND*	K103N (78.4;78.7) E138Q (13.4;7.7)	S	S	S	S	HLR
NWC033	Sanger	ND*	K103N, Y181C	LLR	S	HLR	S	HLR
	NGS	ND*	K103N (84.0;87.9), Y181C (98.0;99.6) K103S (13.0;12.3)	LLR	S	HLR	S	HLR

SSusceptible; PLLRPotential Low-Level Resistance; LLRLow-Level Resistance; IR Intermediate Resistance; HLR High-Level Resistance;*ND: No mutations detected