

**Population and single genome kinetics driving the evolution of
multiple-linked multiclass drug resistance mutations in the viral
protease and reverse transcriptase of HIV-1 subtype C in children
receiving early protease inhibitor based combination therapy**

By Camille Marie Lange

University College London

Thesis submitted for the degree of Doctor of Philosophy

I, Camille Marie Lange confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this had been indicated in the thesis.

Abstract

This thesis examines the evolution of HIV-1 subtype C multiple linked multi-class antiretroviral resistance mutations in the viral protease (PR) and reverse transcriptase (RT) genes of vertically infected children. Emergence of PI resistance on the backdrop of pre-existing non-nucleoside reverse transcriptase inhibitor (NNRTI) resistance could compromise long-term treatment options in such children. We characterised multi-class drug resistance using single genome sequencing (SGS) in children with viraemia while receiving PI-based ART. We applied SGS of HIV-1 protease (PR) and reverse transcriptase (RT) to longitudinal samples from a cohort of the Children with HIV Early Antiretroviral Therapy (CHER) trial with viral loads >1000c/ml after 40 weeks of early ART. Bulk sequencing revealed NVP-selected resistance in 50% of these children while SGS revealed NVP-selected resistance in 70%. Two children had baseline NRTI and PI mutations, suggesting previous maternal ART. Linked multi-class drug resistance following PI-based ART was detected by SGS in 2/10 children. In one child, the majority species contained M184V in RT linked to L10F, M46I/L, I54V and V82A in PR and a triple-class drug resistant variant with these mutations linked to the NNRTI mutation V108I. In the second child, the majority species contained M184V and V82A linked within viral genomes.

I correlated nucleotide variation of PR-RT with the number of single genomes obtained at each time point and ART status and used maximum likelihood trees, recombination analysis, positive selection analysis and co-evolution analysis to describe the evolution of PR-RT of the viral populations. Six children who received early ART for 40/96 weeks only or received continuous ART for the duration of the

CHER trial had clusters of identical sequences from baseline and week 40 of ART. These sequences did not harbour known drug resistance mutations. Therefore one could hypothesize viral replication from a persisting viral reservoir that was established from infection that occurred prior to the initiation of ART. The rooted ML trees of 2 children who developed drug resistance during ART had clusters of identical sequences harbouring common drug resistance mutations from multiple time points which is characteristic of the selection of drug resistant viral populations that cause virological failure during ART. When drug resistant viral populations developed during treatment failure, M184V single mutated viruses were selected from multiple wildtype viral populations but only one population became the major contributor to drug resistant viraemia in both children. Triple-class drug resistant sequences that had common DRMs (M184V, V108I in RT and M46I in PR) did not cluster together. I found no evidence of recombination or coevolving sites in PR-RT for any of these children.

I used a luciferase based single replication cycle assay to examine drug susceptibility and replication capacity (RC) conferred by multi-class drug resistant PR-RT from the 2 children who developed such drug resistant variants. I tested the susceptibility of pseudoviruses to the components of early ART (AZT, 3TC and LPV), the components of second-line therapy for these children (Abacavir (ABC), Didanosine (ddI), Efavirenz (EFV) and (NVP), the PIs Nelfinavir (NFV) and Saquinavir (SQV), which are also approved for use in children and Darunavir (DRV), which has been identified as a PI option needed in paediatric co-formulation. Pseudoviruses with known PI resistance conferring mutations showed reduced susceptibility to all PIs except DRV. Those with known NNRTI resistance conferring mutations showed reduced

susceptibility to EFV and NVP. M184V mutated pseudoviruses conferred high-level resistance to 3TC. In one child, a combination or one of the RT mutations V35T, E36D, T39R, S48T, T165I, K173A, D177E, T200A, Q207D, R211K, V245Q, E248N, D250N, A272P, K277R, E291D, I293V, T296N may be associated with high-level ABC and ddI resistance when genetically linked with M184V.

Population sequence analysis was used to characterize the viral gag genes that encoded matrix, capsid, nucleocapsid, p6, and spacer peptides 1 and 2 along with PR-RT as a single amplicon. I determined the presence of compensatory PI-resistance mutations in gag, drug resistance mutations in PR and RT and other amino acid changes that occurred during ART. To determine the polymorphic nature of these sites, I compared them to a position-specific scoring matrix for gag that was derived from HIV-1 subtype C sequences from children from Sub-Saharan Africa. P453L in the p1/p6 cleavage site of Gag emerged in the viral population of one child during PI-based ART. It was the only amino acid change in Gag that emerged among all children in the study cohort that has been characterised as a compensatory mutation that is selected by and enhances PI-resistance.

This project is the first to identify multi-class drug resistance mutations in PR and RT that were linked on the same genome as well as characterise their development during early PI-based ART in children. Triple class drug resistant viruses detected in the minority species of the viral population of one child demonstrated significant levels of resistance to LPV, SQV, NFV, 3TC and NVP, and established that such variants could compromise future ART regimes if they became the dominant species of the viral population. I note that the small convenience sample (n = 10) chosen for this project limited the power of this study so that findings could not be generalized.

Acknowledgements

I take this opportunity to thank my advisors, Professor Deenan Pillay, Dr. Ravindra Gupta and Dr. Stéphane Hué. I would also like to thank the University College London Grand Challenges in Global Health PhD Programme and the Government of Trinidad and Tobago for funding my research.

Ravi, I would especially like to thank you for seeing me through to the end of this chapter in my life. It would have been impossible for me to achieve this accomplishment without your encouragement, your constant communication, your frank advice and your quick turnover times for reviewing my work. Thank you for encouraging my research and for allowing me to grow as a scientist.

I would also like to thank Dr. Marlen Aasa-Chapman for getting me started with my research project and leaving me with the practical skills necessary to truly optimize my time during the execution of my phenotypic assays. I could not have finished my final experiments without the wealth of knowledge, understanding, reagents and consumables you left with me for these techniques.

Dr. Paul Grant, Ms. Bridget Ferns and Mr. Stuart Kirk from the Clinical Microbiology and Virology Unit of the UCLH NHS Foundation Trust, my deepest gratitude for the countless ways that you assisted me. My laboratory work would have been extremely difficult without your technical help to troubleshoot and complete my SGS assays as well as lending me a workbench in your laboratories for the many times I needed a “clean area”. I would also like to thank Dr. David Bibby at Public Health England for helping me with my cloning techniques and taking the time to answer my many

questions even when there was no answer. You gave me the confidence to continue working in the face of many blank gels and incorrect sequences.

To Dr. Petra Mlcochova, Dr. Katherine Sutherland and soon to be Dr. Sarah Watters of team Gupta: during the good times you were my sources of encouragement and partners in joy; during the bad times, shoulders to cry on without judgment. Thank you for looking at my ugly blank gels and letting me know that the next one will be better. Sarah, in more ways than one - thank you for friendship and I hope to return the favour many times over as you go through 'the struggle'. Petra, you are amazing and a mentor in life and science!

I could not have achieved this goal without the support of my friends and family. To my father, thank you for providing me with so many opportunities to achieve all I strived for, I could not ask for more. To my mother, thank you for supporting me in the last few months of writing this thesis: your pride in me helped me to re-discover my pride in myself. To my siblings, Jean-Paul, James and Siân thank you for listening to my rants about anything and everything even when you did not understand what I was talking about and yet supported me despite. To my grandmother: I appreciated your support through this process. Danille, my best friend and partner in success all these years: "May we continue to achieve all our goals and strive for our best."

Finally, Mario, thank you for knowing my potential and continuing to push me towards it despite my reluctance, despite my bouts of no confidence, and despite the many times I did not know how to overcome barriers to success.

Table of Contents

Chapter 1	Introduction	13
1.1	HIV etiology and diversity.....	13
1.1.1	HIV epidemic.....	17
1.1.2	Prevention of mother-to-child transmission of HIV	18
1.1.3	Course of HIV-1 infection	21
1.1.4	Drug resistance in children	28
1.1.5	Significance of drug resistance in the minority sub-populations.....	32
1.1.6	Approaches for measuring population diversity	34
1.2	HIV-1 structure and replication	36
1.2.1	The HIV-1 genome	36
1.2.2	HIV-1 virion.....	36
1.2.3	HIV-1 replication cycle.....	39
1.3	HIV-1 Antiretrovirals.....	50
1.3.1	Targets and mechanisms of action of HIV-1 antiretrovirals.....	50

1.4	Antiretroviral resistance.....	53
1.4.1	Mechanisms of HIV-1 genetic diversity	53
1.4.2	Reverse transcriptase inhibitor resistance	54
1.4.3	Protease inhibitor resistance	57
1.5	Thesis rational.....	62
Chapter 2	General Materials and Methods	67
2.1	Clinical samples.....	67
2.2	Molecular biology techniques.....	68
2.2.1	Viral RNA extraction.....	68
2.2.2	Oligonucleotide primer design.....	70
2.2.3	cDNA synthesis	71
2.2.4	Endpoint dilution PCR.....	76
2.2.5	Sanger sequencing	79
2.2.6	Sequence analysis and multiple sequence alignments.....	79
2.2.7	Subtype and drug resistance determinations.....	82

2.2.8	DNA quantification.....	83
2.2.9	Addition of restriction endonuclease sites to amplicons by PCR	83
2.2.10	Cloning into p8MJ4-HSC	84
2.2.11	Plasmid preparation	84
2.3	Phylogenetic and bioinformatics analyses	89
2.3.1	Phylogenetic reconstruction of PR-RT sequences.....	89
2.3.2	Assessment of recombination in PR-RT.....	91
2.3.3	Intra-patient analysis of selection pressure on the HIV-1 PR and RT genes	91
2.3.4	Co-evolution analysis.....	92
2.3.5	Position Specific Scoring Matrix	92
2.4	Tissue culture	93
2.4.1	Pseudovirus production.....	93
2.4.2	FuGENE® 6 transfection for pseudovirus production	93
2.4.3	Single-replication cycle drug susceptibility assay	94
2.4.4	Replication capacity.....	95

2.4.5	Protease inhibitor susceptibility assay	95
2.4.6	Reverse transcriptase inhibitor susceptibility assay.....	98
Chapter 3 Results		100
3.1	Single genome analysis reveals linked, multi-class drug resistance in HIV-1 infected children from the Children with HIV Early Antiretroviral (CHER) study 100	
3.1.1	Results.....	101
3.1.2	Discussion.....	117
3.1.3	Conclusions.....	121
3.2	Phylogenetic analysis of PR-RT to characterise the development of multiple linked multi-class drug resistance	123
3.2.1	Results.....	126
3.2.2	Discussion.....	150
3.2.3	Conclusions.....	155
3.3	Longitudinal trends in replication capacity and drug susceptibility associated with development of multiple linked multi-class drug resistance and characteristics of Gag during PI-based ART	156

3.3.1	Results.....	158
3.3.2	Discussion.....	176
3.3.3	Conclusions.....	181
Chapter 4	Final Discussion, Project Limitations and Future Work.....	182
Chapter 5	Appendices.....	188
5.1	Appendix A.....	188
5.2	Appendix B.....	189
5.3	Appendix C.....	192

Chapter 1 Introduction

1.1 HIV etiology and diversity

In 1982 the Centre for Disease Control and Prevention was the first to define Acquired Immune Deficiency Syndrome (AIDS) [1] as a disease predictive of defective cell-mediated immunity without a known cause for disease susceptibility. Kaposi's sarcoma, *Pneumocystis carinii pneumonia* and other opportunistic infections were predictive of this cellular immunodeficiency [1]. A year later the causative agent of AIDS, Human Immune Deficiency Virus (HIV), was isolated at the Institute de Pasteur [2] and then the National Cancer Institute [3] and identified as a distinct retrovirus from the family of Human T-cell Leukemia viruses (HTLV). HIV became known by two names, Human T-cell Lymphotropic virus type 3 (HTLV-III) and Lymphadenopathy associated virus (LAV). In 1986 a subcommittee of the International Committee on the Taxonomy of viruses suggested the name HIV for this newly described etiological agent of AIDS and characterized it as a member of the Lentiviridae family of retroviruses, which could infect a range of species including cats, horses, equine species and primates [4].

The relatedness of HIV genomic sequences determined the origin of HIV as a zoonosis of Simian Immunodeficiency Viruses (SIV) from primates to humans [5]. Zoonoses were previously observed with viruses like HTLV-1 and smallpox [5]. The HIV epidemic was divided into two types: HIV-1 and HIV-2. In 1999 a strain of SIV retrieved from the *Pan troglodytes troglodytes* subspecies of chimpanzees found in Cameroon was reported as the origin of at least three independent introductions of SIVcpz and was determined as the primary reservoir for HIV-1 [6].

There are four known strains of HIV-1 of which the main strain, group M, leads the worldwide epidemic [7]. The “non-M, non-O strain” called group N is rare and has not been detected outside of Cameroon [8]. The outlier group O is also most frequently detected in Cameroon as 2% of all HIV infections [9]. Group P was identified in 2009 from a Cameroonian woman living in France and this strain was determined to be more closely related to gorilla SIV (SIV_{gor}) than SIV_{cpz} [10].

The evolutionary flexibility of HIV has allowed different strains to be established around the world. Single strains of HIV can populate particular geographic locations so that a population is established with time. Sequence analyses of viruses from these populations demonstrate greater sequence similarities within them, compared to those from other geographic locations. Given the effect of these founder infections, HIV-1 group M is sub-divided into subtypes designated A-D, F-H, J, K (no E and no I) and circulating recombinant forms, (CRFs). CRFs are derived from recombinations of genomes from two or more Group M subtypes. Subtype A and its recombinant forms dominate the epidemic in Central and South Asia and subtype B dominates the epidemic in the Americas, Europe and Australia [11]. Subtype C and its related CRFs are the most prevalent worldwide, making up >50% of global infections [11]. This is mainly due to the pandemic levels of subtype C in highly HIV-1 positive populous countries in Southern Africa [12], South America [13] and China [14, 15]. The geographical distribution of HIV-1 subtypes worldwide is shown in Figure 1.1

Similar to HIV-1, HIV-2 was also determined as a zoonosis. The primary reservoir was identified as a SIV prevalent in a subgroup of West African sooty mangabeys

called *Cercocebus atys* (SIVsm) [6, 16]. Eight subtypes of HIV-2 have been identified. These are groups A through H, of which groups A and B constitute the HIV-2 epidemic. Groups A and B are primarily endemic in West Africa. Group A has a prevalence rate >1% in Cape Verde Islands, Côte d'Ivoire, The Gambia, Ghana, Guinea-Bissau, Mali, Senegal, Mauritania, Nigeria, and Sierra Leone as well as Angola, Mozambique [17] [18] and reported in Brazil, India, Europe and the United States of America at a much lower prevalence [18]. Group B has been confined to West Africa and reported mainly in Côte d'Ivoire and Ghana [17]. Groups C through H have been identified in six people from West Africa as single independent transmission events from SIVsm in Liberia [19], Sierra Leone [19, 20] and Côte d'Ivoire [21, 22]. Group A was estimated as 1940 ± 16 years and group B as 1945 ± 14 years [23].

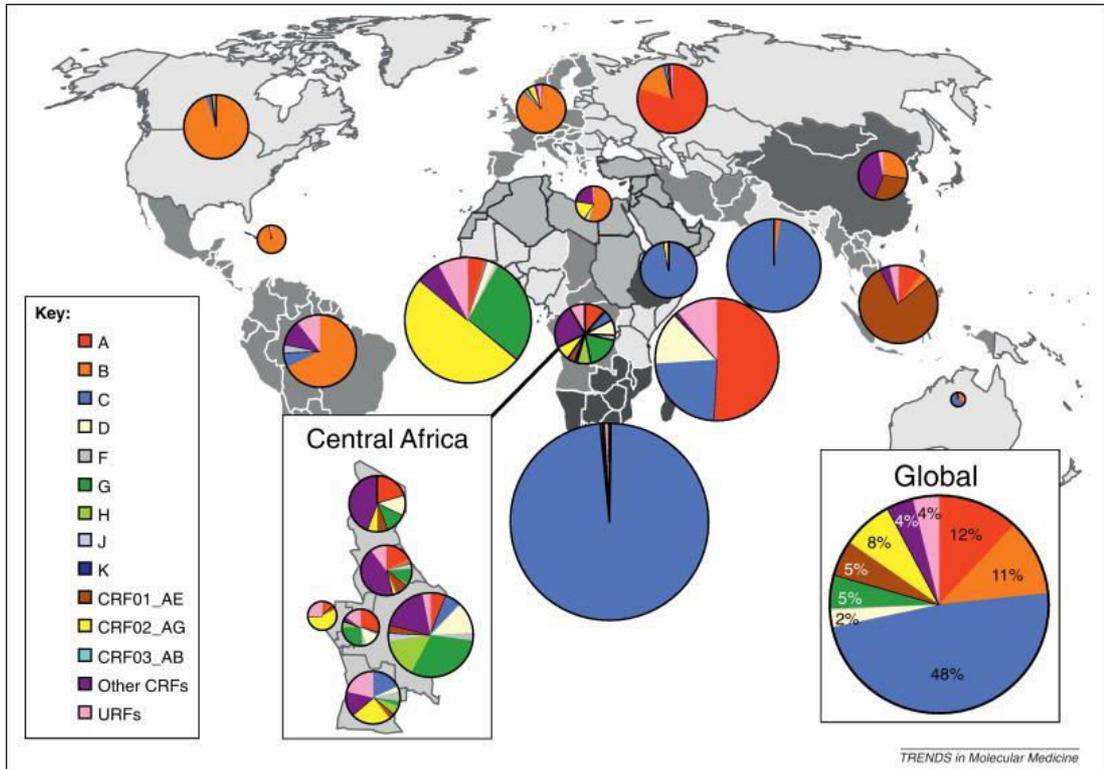


Figure 1.1 Geographical distribution of HIV-1 subtypes. The pie charts depict the proportion of each subtype or CRF in each geographical region from 2004 to 2007 and are colour-coded according to the legend on the left side of the figure. The HIV-1 subtype distributions found around the world and within Central African countries are shown in the insets of the main figure. This figure was taken from Hemelaar et al [24] and was originally reproduced and adapted with permission from [11].

1.1.1 HIV epidemic

1.1.1.1 The adult HIV epidemic

As of 2011 the number of people living with HIV worldwide was estimated at 34.0 million and 58% of adults who required ART received it. Sub-Saharan Africa bears the highest burden of HIV infections with 69% (23.5 million) of the global estimate living with HIV and as ART coverage improved, 55% of adults who needed ART received it in 2011 [25]. Improvements in HIV prevention strategies have also seen the number of new infections in this region decrease from 2.4 million in 2001 to 1.8 million in 2011 and the improved accessibility to antiretroviral therapy (ART) has seen the number of people living with HIV increase from 20.9 million to 23.5 million during the same time [25].

1.1.1.2 The paediatric HIV epidemic

The first reports of AIDS in children were in 1985 and the disease was described as rapidly progressing with high mortality rates for HIV infected children [26]. Between 2009 and 2013, there was a 43% reduction in the number of new infections in children and between 2011 and 2013, the global estimate of children living with HIV dropped from 3.4 million to 3.2 million [27-29]. Similar to the adult epidemic, sub-Saharan Africa bears the highest burden (91%) of children living with HIV. A large gap also exists between paediatric and adult treatment coverage in this region with only 24% of children who needed ART receiving it [25].

1.1.2 Prevention of mother-to-child transmission of HIV

Within the last 15 years there have been continuous improvements to prophylactic strategies for the prevention of mother-to-child transmission (PMTCT) of HIV-1 (Table 1). In resource-rich countries, where these strategies were sufficiently implemented, MTCT decreased to $\leq 2\%$ [30]. Without prophylaxis, the MTCT for children <15 years of age was estimated as 12% - 45% [31, 32]. Without breastfeeding this rate improved to 15 - 30% compared to 30 - 45% with breastfeeding [32]. It is estimated that 35% of MTCT occurs during pregnancy, 65% as perinatal infections and 7% - 22% from breastfeeding [33, 34]. In 2001 the United Nations committed to reducing paediatric HIV infections by 50% by 2010 and focused heavily on the scale-up of country PMTCT programmes [35].

Single-dose NVP (sdNVP) as MTCT prophylaxis in infants (2mg per kg orally within 72 hours of birth) has been shown to reduce vertical transmission rates by 36% compared to the absence of PMTCT regimes [36]. When sdNVP is used in combination with maternal administration of single dose NVP (200 mg orally at onset of labour), MTCT rates are reduced even further [37, 38]. Single-dose NVP has been the most common PMTCT regimen used in resource-limited settings and has been a WHO recommendation since 2001 [39]. It is not the most effective prophylaxis regime, however its operational effectiveness for PMTCT, especially in resource-poor settings makes it a preferred treatment option [40]. As of 2010 the WHO no longer recommended sdNVP for PMTCT because longer-term NVP use was determined as more effective for this purpose, especially with prolonged breastfeeding [41-43] (Table 1). The WHO now recommends daily doses of NVP

from birth and until 1 week after the end of the breastfeeding period or for 4-6 weeks if breastfeeding ends before the first 6 weeks of life [44-46] (Table 1).

The WHO reported PMTCT therapy coverage as drastically increased from 9% in 2004 to 48% in 2010. With increased access to paediatric treatment service, there has been a 29% decrease in AIDS-related deaths in children between 2005 and 2011 [27, 45, 47]. From 2009 to 2011, PMTCT interventions prevented ~ 409 000 vertical transmissions of HIV in low- and middle-income countries [27, 47] and there was a 24% reduction in the number of new infections between 2009 and 2011 [46].

Recommendation name	Year of WHO recommendation	Recommendation		Key Trial / Study
		Mother	Child	
Option B+	2013	Triple ART throughout pregnancy and then for life, despite clinical status or CD4 count	Same as Option A	HPTN052 [48]
Option B	2010	<ul style="list-style-type: none"> ➤ Triple ART throughout pregnancy and 7 days post breastfeeding. ➤ Triple therapy continued for life if CD4 count \leq 500 cells/mm³ 	Same as Option A	SWEN [41] Shapiro et al [49]
Option A	2006	<ul style="list-style-type: none"> ➤ AZT from week 28 of gestation ➤ sdNVP at onset of labour ➤ 7 days of postnatal AZT/3TC 	<ul style="list-style-type: none"> ➤ BF: sdNVP from birth and then for 6 weeks ➤ No BF: once-daily NVP or twice-daily AZT from birth for 4-6 weeks 	Mashi [50] ANRS1201/1202 DITRAME PLUS [51]
	2004	<ul style="list-style-type: none"> ➤ AZT from week 28 of gestation ➤ sdNVP at onset of labour 	sdNVP at birth	HIVNET 012 [38]
	2001 Before 2001	sdNVP at the onset of labour AZT from week 14 or 36 of gestation	sdNVP at birth 6 weeks of postnatal AZT (optional)	HIVNET 012 [38] ACTG 076 [52, 53]

Table 1. Timeline of WHO recommendations for PMTCT of HIV-1. BF = breast-feeding.

1.1.3 Course of HIV-1 infection

90% of worldwide HIV infections are a result of oral, rectal, or vaginal mucosa exposures during sexual intercourse [54-57]. It can also be transmitted by inoculation with contaminated blood or blood products, for example needle sharing during intravenous drug use and needle stick injuries, blood transfusions [58] and vertical transmission routes in infants from mother-to-child, i.e. in utero, during delivery and through breast milk [32, 33, 59].

1.1.3.1 Course of HIV-1 infection in adults

In adults the course of infection in typical HIV-1 progressors, can be divided into 3 phases: 1) primary infection 2) clinical latency and 3) AIDS [60]. During primary infection the viral load can be as high as 10^8 HIV-1 RNA copies / millilitre of plasma (copies/ml) and coincide with a decline in CD4+ T lymphocyte counts (CD4 counts). This is followed by a reduction in viraemia, thought to be a result of the development of a cytotoxic CD8+ T lymphocyte (CTL) immune response to the infection. Clinical latency follows and can last for up to 10 years. During this time there is a steady loss of CD4+ T cells that correlates with the viral load. During this time, these cells continue to be infected, eradicated and regenerated. During AIDS the CD4 count declines rapidly and in tandem with an increase in viral load. It is at this stage that patients succumb to opportunistic infections associated with late stage disease. The course of HIV-1 infection in normal progressors is shown in Figure 1.2.

1.1.3.2 Course of HIV-1 infection in infants

The course of HIV infection in infants can have two profiles [61-65]: the first profile follows the same course described for normal adult HIV-1 progressors (Figure 1.2) where HIV progresses to AIDS between 8 and 10 years after initial infection. The second profile is characterized by a more rapid progression to AIDS with increased risks of opportunistic infections and mortality. In vertically infected infants, rapid progression to AIDS is usually characterized by a positive HIV DNA polymerase chain reaction test within 48 hours of delivery and $\leq 30\%$ CD4+ T-cells out of total lymphocytes within the first week of life and may be a sign of *in utero* transmission of HIV (early HIV infection) [66]. Increasing viral loads with decreasing CD4+ lymphocyte counts is associated with increased risks in mortality in all HIV-1 infected children [67].

There are several factors that help to predict rapid progression to AIDS in infants, but it is important to note that none of these factors can fully predict this. A prospective study published in 1996 [68] determined that risk for developing AIDS, defined as Centers for Disease Control and Prevention classification Category C by 12 months old in vertically infected children were: 1) liver and spleen enlargement or adenopathies in infected newborns, 2) $< 30\%$ CD4+ cells at birth, also seen in other paediatric cohorts [62, 69] and 3) positive HIV antigen testing at birth. The authors determined that these parameters “were strongly interrelated and could reflect active disease onset in utero in some cases of early, severe HIV-1 disease in childhood” [68].

Furthermore, Dickover et al [70, 71], Kuhn et al [72] and Rich et al [73] showed that vertical infections of HIV-1 that occurred in utero are associated with rapid progression to AIDS in these children. In addition, Chen *et al* [74] showed that the presence of the MHC class I allele HLA-A*2301 has also been associated with rapid progression to AIDS and death by 2 years of age for vertically infected children. Two studies [69, 75] have also found that maternal AZT treatment was associated with rapid disease progression in vertically infected children, an effect which is yet to be understood.

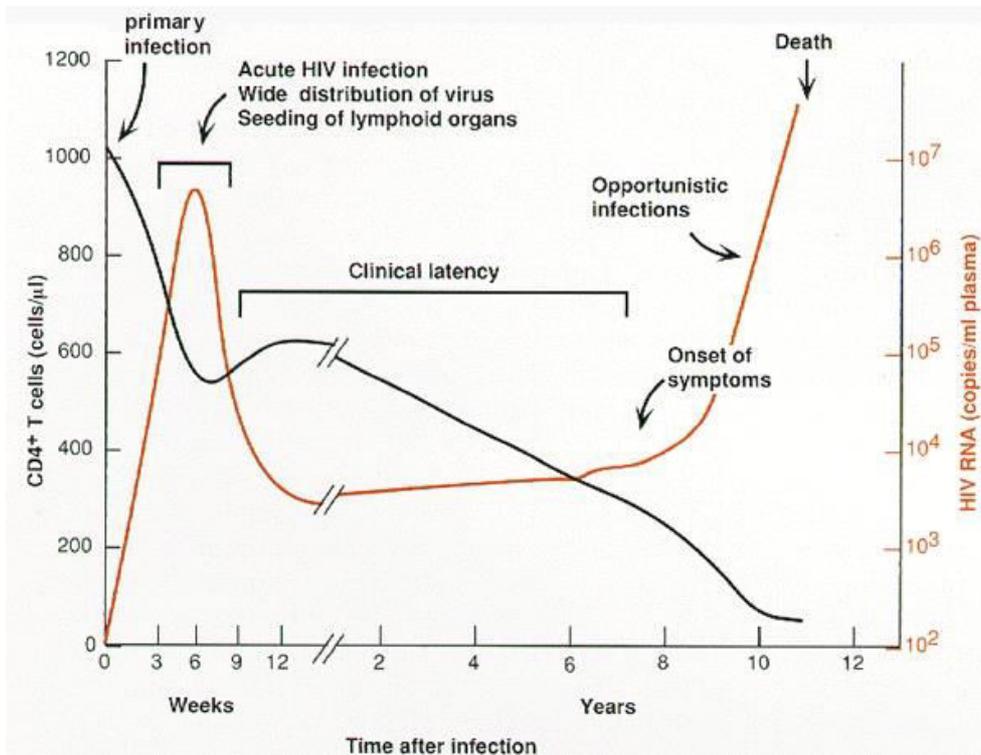


Figure 1.2 Course of HIV-1 infection. The graph shows the typical course of infection in HIV-1 normal progressors. Rates and values of CD4+ T-lymphocyte loss and virus load increases vary between patients. This figure was taken from Coffin *et al* [60].

1.1.3.3 Treatment of HIV in children

Without ART, 50% of HIV-infected children did not live past 2 years of age and those who survived face a severely reduced quality of life associated with AIDS-related disease and failure to thrive; 80% of these children die by age 5 [76]. In 2004 Luzuriaga et al [77] showed that children who received ART within the first 12 weeks of life had a better virological outcome (viral load <400 copies/ml) by week 48 of ART compared to children who received ART after the first 3 months of life. In 2007, preliminary results of the Children on Early Antiretrovirals (CHER) trial showed that child mortality and disease progression in HIV-1 positive children improved by ~75% when PI-based ART was started within the first 12 weeks of life [78] compared to previous treatment guidelines to start ART when immunosuppression was detected [79, 80]. These results influenced the World Health Organization (WHO) to change its treatment guidelines for HIV-infected infants so that in April 2008, first initiation of ART was strongly recommended for in children <1 year old immediately after a positive HIV diagnosis, despite their state of health [81].

Antiretroviral formulations for young children are dependent on the age and weight of the child and may also require multiple dosing since under-dosing carries the risk of the development of drug resistance, particularly in resource-poor settings [82]. Since most children with HIV infections live in resource-poor settings, they also have the highest risk of developing drug resistance [82]. A meta-analysis by Sigaloff et al [83] quantified the prevalence of drug resistance in this context for children who failed first-line ART at 90% (95% CI 88–93): 80% harboured NRTI resistant viruses, 88% harboured NNRTI resistance and 54% harboured PI resistance. They also

concluded that shortages of appropriate antiretroviral co-formulations or incorrect dosage calculations were the major sources of under-dosing [83] (the majority of the studies in their meta-analysis were published after 2007 when WHO paediatric treatment guidelines were changed to recommend early initiation of first-line ART for infants and children [81]).

The recommended drug dosages for children are intended to achieve comparable concentrations to those seen in adults with a good virological response to ART, and work continues to determine optimal dosages for infants and children; drug toxicity also limits the antiretroviral options that can be used in children [28, 79, 81, 83-86]. Of the 28 antiretrovirals currently approved for use in HIV treatment and care, only 5 are recommended for use in children of any age and they belong to only 3 of the 6 drug classes available. Even less are approved for use in infants because of drug toxicity, poor virological response or insufficient data to support their use (Table 2).

Drug Class	Antiretroviral (Abbreviation)	Liquid formulation (Y/N)*	Age group dosing
Nucleos(t)ide reverse transcriptase inhibitors (NRTIs)	Zidovudine (AZT)	Y	All age groups
	Lamivudine (3TC)	Y	All age groups
	Stavudine (d4T)	Y	All age groups
	Didanosine (ddI)	Y	≥2 weeks old
	Abacavir (ABC)	Y	>3 months old
	Emtricitabine (FTC)	Y	≥3 months old
	Tenofovir (TDF)	N	>12 years old (only with Hepatitis B coinfection)
Non-nucleoside reverse transcriptase inhibitors (NNRTIs)	Nevirapine (NVP)	Y	All age groups
	Efavirenz (EFV)	Y	>3 years old and >10kg
	Etravirine (ETR)	N	≥6 years old
Protease inhibitors (PIs)	Ritonavir (RTV)**	Y	All age groups
	Lopinavir/Ritonavir (LPV/r)	Y	≥2 weeks old
	Nelfinavir (NFV)	N	>2 years old
	Saquinavir (SQV)	N	≥16 years old and >25kg
	Atazanavir (ATV)***	N	100mg capsules discontinued in the last year, therefore dosing recommendations are being re-evaluated
	Darunavir (DRV)***	N	>3 years old and 10kg (efficacy and co-formulation evaluations have not been conducted)
Integrase Strand Transfer Inhibitors	Dolutegravir***	N	≥16 years old
	Raltegravir***	Y	>4 weeks old and >3kg

Table 2. Antiretrovirals recommended in children according to FDA and WHO guidelines. *Y/N=Yes/No. **RTV is usually used as a pharmacokinetic enhancer in dual PI regimens. *** Important notes or newly approved for use in children. This table is an adaptation of the 2013 WHO report on Antiretrovirals in Paediatric patients: Focus on Young Children [87] and Annex E of the 2009 report on Antiretroviral Therapy got HIV Infection in Infants and Children: Towards Universal Access [28].

1.1.4 Drug resistance in children

HIV positive children require long-term use of ART from childhood to adolescence to adulthood, therefore they especially require first-line therapy to be highly effective and durable by minimizing the emergence of drug resistance and drug toxicity.

Similar to adults, drug resistance in children is associated with poor adherence and suboptimal treatment regimes. Children may be more prone to developing drug resistance because they are kept on failing ART regimes for a longer time compared to adults and the risk of developing drug resistance increases when the same ART regime is continued with a detectable viral load. Reasons for continuing a failing therapy regime in children include limited treatment options, dependence on carers and deficient paediatric ART formulations [82].

In 2011, Sigaloff et al [83] reported that first-line therapy failure in the developing world was associated with at least one detectable drug resistance mutation in 90% of HIV positive children. Before the overwhelming success of PMTCT programmes in the developed world, children were two times more likely to fail triple therapy than adults and this risk increased with the duration of therapy into adolescence.

In low- and middle-income countries, PMTCT programmes and regimes have a significant impact on drug resistance in vertically infected children when the intervention is unsuccessful. Drug resistance may be a result of mother-to-child transmission of drug resistant viruses or peripartum selection of drug resistant viruses in the child by maternal antiretrovirals or PMTCT interventions in the child.

In 2007, Arrive et al [88] reported 52.6% of children who were exposed to prophylactic single-dose NVP developed detectable NNRTI resistance and Y181C

was most frequently detected. In 2010, Palumbo et al [89] showed that a PI-based first line treatment regime was more appropriate for vertically infected children who received prophylactic NVP because NNRTI resistance selected by this intervention was predictive of virological failure on NVP-based first-line regimes in children. As a result, all global paediatric guidelines were amended to recommend PI-based therapy in children with prior NVP exposure or for those who have a high risk of developing NNRTI resistance. Furthermore, the 2012 Nevirapine Resistance (NEVEREST) study [90] determined that a higher risk of virological failure was still associated with NNRTI-based ART despite initial suppression of the viral load using a PI-based regime and children with detectable baseline NNRTI resistance were also more likely to fail NVP-based ART by 3 years of age.

NVP-based first-line therapy is still recommended for children without prior NNRTI exposure, but this recommendation remains debatable and the literature reports contrasting views on this topic. Many studies have reported an increased risk of NNRTI-based therapy failure in children <3 years old despite no evidence of prior NNRTI exposure [91-94], but other studies report no differences in treatment outcomes between NNRTI and PI-based regimes in such children [95].

Perinatally infected children enter adolescence with increased risks of treatment failure because of the development of multiclass drug resistance because of antiretroviral exposure from failed MTCT prophylaxis at birth and sub-optimal combination antiretroviral therapy throughout childhood. There is also limited range and formulations of drugs approved for paediatric use and drug toxicity. In Europe and North America, studies showed that perinatally infected adolescents were

exposed to five [96] to eight [97] antiretroviral regimes by the time they were transferred to adult care and one particular European study found that 20% of children <16 years old were failing cART with triple class drug resistance after 8 years of therapy [98]. Perinatally infected children in the developing world face even higher risks of developing multi-class drug resistance because treatment failure is not detected rapidly enough to prevent the accumulation of drug resistance mutations.

Some studies report that a small proportion of the research cohort had virological failure during first line therapy in the absence of known drug resistance mutations. For example Hamers *et al* [99] found 29.6% (n=166) of patients from 6 countries in Sub-Saharan Africa fit this criteria, Odaibo *et al* [100] also found 30% (n=46) of patients from a Nigerian cohort found the same and Murphy *et al* [101] also found 13.5% (n=141) of patients from a South African cohort were also failing first line therapy without detecting drug resistance mutations.

1.1.4.1 HIV-1 subtype and ARV susceptibility

The susceptibility of HIV-1 subtype B and drug resistance-conferring mutations in this subtype is well studied, however differences in ARV susceptibility of the other subtypes compared to subtype B is still to be defined. Genetic variation among HIV-1 subtypes has the potential to yield distinct drug resistance mutation profiles under the selection pressure of antiretrovirals. A testament to this is that popular HIV-1 drug resistance databases like the HIV drug resistance database run by Stanford University [102] include drug resistance mutation frequencies for non-subtype B *pol* sequences, but mutations are defined as differences compared to a consensus B

reference sequence. There is an 8%-10% difference in genetic variation among HIV-1 subtypes [103].

Discussion continues about the effect of natural amino acid background of HIV-1 subtypes on the magnitude of ARV susceptibility conferred by mutations that are already known to confer antiretroviral drug resistance and predispositions to the development of drug resistance [104, 105]. For example NVP or EFV is more likely to select V106M in HIV-1 RT in subtype C viruses, whereas V106A is more commonly selected in subtype B. This is because of a silent polymorphism in the codon, commonly seen in subtype C that requires 1 nucleotide change to become a resistance conferring mutation (GTG→ATG) [106, 107]. Subtype B generally requires 2 nucleotide changes for the same mutation to develop (GTA→ATG). Another example is the D30N mutations in HIV-1 PR. When this mutation is selected by NFV in subtype B, it confers resistance to this PI by making the flap region of protease more flexible in the PI-enzyme complex [108, 109]. However the same effect is achieved in subtype C proteases when D30N and N83T are both selected in the enzyme [108, 109].

1.1.4.2 The CHER trial

The CHER trial began in 2005 and its aim was to determine if there was an optimal time to start ART in vertically infected children. The research team hypothesized that early ART administered until the 1st or 2nd birthday of the child may delay disease progression and delay the need for long-term, continuous ART [110]. Prior to the CHER trial, there were no comparative studies to inform ART guidelines for such children, and infants had poor disease outcomes and high mortality rates [76]. The

inclusion criteria for the CHER trial were (1) <12 weeks old (2) a positive DNA PCR test (3) ART naïve with the exception of PMTCT NVP (4) CD4% > 25%. 377 participants were enrolled in the study based on this criteria [78]. Participants were randomly assigned to one of three treatment arms: Arm 1 was deferred ART [84], Arm 2 was early ART until week 40; Arm 3 was early ART until week 96.

The primary endpoint of the study was death or virological failure (viral load >1,000 copies/ml) during early ART [78]. The secondary endpoints of the study [78, 110, 111] were a cumulative risk of disease progression that led to hospitalisation, the occurrence of Grade 3 or 4 adverse events as defined by the United States National Institute of Allergy and Infectious Diseases [112] and the development of drug resistance. Children were followed-up for a minimum of 3.5 years. Preliminary results revealed that starting ART before 12 weeks of age (Arms 2 & 3) reduced infant mortality rates by 76% compared to deferred therapy (Arm 1). Early ART was also associated with a 75% improvement in disease progression compared to deferred therapy. The frequency of breastfeeding in each Arm of the Study was 20% [110].

1.1.5 Significance of drug resistance in the minority sub-populations

Several studies have shown that the drug resistant minority variants in the viral population that are not detected by the standard population based genotyping techniques can be revealed with more sensitive approaches [113-116]. The clinical implications of these minority species that develop in the viral populations of HIV-1 infected patients continue to be discussed. They have been detected in women and mothers [117-120] and children [119-121] who were treated with sdNVP for

PMTCT, acute seroconverters [122], patients receiving early therapy [123], those in structured treatment interruption programmes [124] and patients experiencing virological failure [116, 123, 125-128]. Some studies have shown that drug resistant minority species can develop into the major viral population during first-line therapy [127], during salvage therapy [116, 129-131] and in, treatment-experienced patients [127, 132]. A 2011 systematic review and pooled analysis by Li et al [127] found that the detection of minority resistance before starting ART was associated with a 2.5- to 3- fold risk of virological failure during first-line ART and this increased risk of virological failure was associated with NNRTI resistance. More recently, Cozzi-Lepri et al [133] found that minority RTI resistance that was detected before starting first-line RTI based ART was associated with >2-fold risk of virological failure during first-line ART. However the low-level drug resistance mutations detected at baseline were not always the same mutations detected with virological failure during therapy, a finding that is not uncommon in the literature and continues to be a topic of discussion for researchers. Also unlike Li et al's findings, Cozzi-Lepri et al found that the risk of virological failure increased when there was a higher mutation load, i.e. a higher number of pre-existing minority drug resistance variants in the viral population: Patients with 400 to 1000 minority drug-resistant HIV-1 variant copies/ml plasma had an intermediate risk of virological failure on first therapy whereas those with >1000 minority drug-resistant HIV-1 variant copies/ml plasma had a high risk.

1.1.6 Approaches for measuring population diversity

Commercial and ‘home-brew’ population-based sequencing tends to characterise the dominant viral species in the plasma sample. The sensitivity of these assays has been reported as $\geq 20\%$, which does not reflect the true diversity of the viral population and does not always provide complete information on the genetic linkage of drug resistance mutations (DRMs) on the same genome [134]. Multiple methods have been developed to characterize the viral population, each with its limitations. There are highly sensitive point mutation assays such as the oligonucleotide ligation assay, parallel allele-specific sequencing and allele-specific real-time PCR. Next generation sequencing (NGS) is also a highly sensitive approach with many of the advantages of the previously mentioned approaches and is also often used for whole genome reconstruction. The sensitivity range of these methods to measure HIV-1 population diversity has been reported at 0.1-1% [113, 135, 136]. Some clinical limitations of these methods are that point mutation assays can be limited by insensitivity of primers for their templates and analyzing thousands of sequences to confirm a rare variant [113]. NGS has a very high throughput of sequences, requiring specialist bioinformatics expertise to be executed as well as to manage the data and analyse and interpret results. Read lengths are also shorter in the case of 454™, Illumina®, and Ion Torrent™ platforms [137] compared to Sanger sequencing, limiting the power to determine the evolution of genetically linked drug resistance mutations [135]. In the future, longer read lengths with this technology will no doubt become available.

Another approach is single genome sequencing (SGS), where viral RNA is reverse transcribed and the complementary DNA (cDNA) product is diluted so that one cDNA molecule is used for PCR amplification. This was followed by sequencing of a defined region of the viral genome. This is a highly sensitive method as long as a sufficient number of single genomes are analysed [138, 139] and the starting concentration of cDNA for the dilution series is sufficient to minimize re-sampling of the same viral genome [134, 138, 139]. Another approach is to clone the products of one-step reverse transcription-polymerase chain reaction amplification (RT-PCR) of a region of interest in the viral genome and then randomly select clones for sequencing. This may underestimate the diversity of the viral population because of poorly reactive RT-PCR primers for their templates. This approach also has a higher risk of producing artefactual mutations because of PCR-mediated recombination between viral templates, but nonetheless has been shown to be equally effective at measuring population diversity compared to SGS [114]. For this project, SGS was used because of the reduced risk of producing artefactual mutations compared to cloning. SGS was also chosen over NGS techniques because it could detect genetic linkage along longer sequence.

1.2 HIV-1 structure and replication

1.2.1 The HIV-1 genome

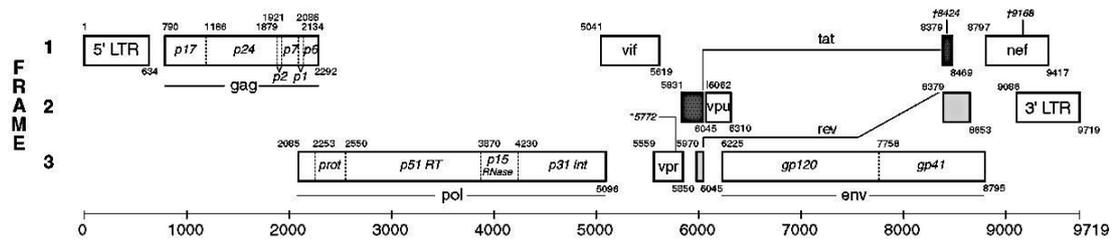
The HIV-1 genome is roughly 9.7kb (Figure 1.3A) and exists as diploid, linear, positive sense single-stranded RNA within the virion and double-stranded proviral DNA after it is reverse transcribed in an infected cell. The HIV-1 genome encodes the structural (*gag* and *env*) and enzymatic (*pol*) proteins as well as proteins which regulate transcription and transport of viral RNA. The structural elements encoded by *gag* are the viral matrix (MA/p17), capsid (CA/p24), nucleocapsid (NC/p7) and p6 and the viral envelope is encoded by *env*, detailed in Section 1.4.3.2. The enzymatic proteins encoded by *pol* are the viral protease (prot/PR) detailed in Section 1.2.4 and 1.2.4.1, reverse transcriptase (p51/RT, RNase H (p15) detailed in Section 1.2.3 and 1.2.3.1 and integrase (INT/p31) detailed in Section 1.2.5 and 1.2.5.1. The regulatory proteins are: transactivator of transcription (Tat), regulation of expression of virion proteins (Rev), negative regulatory factor (Nef), viral infectivity factor (Vif), viral protein unique to HIV (Vpu) and viral protein R (Vpr).

1.2.2 HIV-1 virion

HIV is a lentivirus, which is a genus of the Retroviridae family. Lentiviruses are roughly 80-130 nanometers in size, enveloped and their genetic material is single stranded RNA (ssRNA) encased in an icosahedral capsid [140]. The HIV-1 virion (Figure 1.3B) is approximately 110 nanometers. Its outer envelope consists of a lipid bilayer derived from host cells as well as the viral envelope glycoproteins. Within the envelope is a 'matrix' shell made up of viral matrix proteins, which are anchored to the envelope by N-terminal myristoyl groups [141]. Matrix encapsulates the viral

capsid that encased the genetic material of the virus. Gag, Pol and Env proteins are initially translated within precursor polyproteins. The Gag precursor Pr55^{Gag} (55kDa) is cleaved by the viral protease to produce mature MA, CA, NC and p6 [142]. The GagPol precursor Pr160^{Gag-Pol} (160kDa) is also processed by the viral protease to produce the mature Gag proteins and the *pol* encoded enzymes PR, RT, RNase H and IN [142]. The Env precursor gp160 (160kDa) is processed by host cell proteases into the mature transmembrane glycoprotein gp41 and the surface glycoprotein gp120 [143]. The ends of the viral RNA have terminal direct repeats (R-regions). Downstream of the 5' R-region is the U5 region (5' unique regulator sequence) and upstream of the 3'R region is the U3 region (3' unique regulatory sequence). On both end of proviral DNA are long terminal repeats (LTRs) containing U3-R-U5 sequences [144] (Figure 1.4).

(A)



(B)

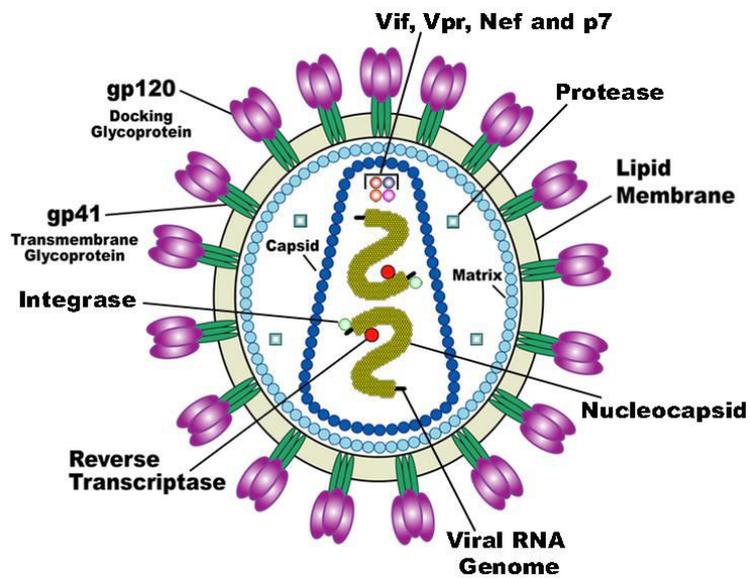


Figure 1.3 (A) Organization of the HIV-1 genome/gene map (~9.7kb). The nucleotide position numbering is relative to HXB2CG (accession number K03455). Rectangles represent open reading frames. Shaded rectangles represent the spliced exons of *tat* and *rev*. Numbers in the upper left corner of each rectangle indicate the start of each gene. Numbers in the lower right corner of each rectangle indicate the last position of the stop codon. The start of *pol* is from the first “T” of the stem loop TTTTTTTAG sequence for ribosomal slippage for the -1 frameshift for Gag-Pol polyprotein translation [145]. (B) Organisation of the mature HIV-1 virion [146].

1.2.3 HIV-1 replication cycle

There are two phases of the HIV replication cycle (Figure 1.8): (1) early and (2) late. The early phase includes binding of the virus to the cell surface receptors, fusion of the viral and cell envelopes, viral entry, reverse transcription of the viral genome and integration of the viral genome into host cell DNA. The late phase of the cycle includes regulated expression of integrated provirus, budding and maturation of the virion.

1.2.3.1 Early phase

1.2.3.1.1 Binding, fusion and entry of HIV into the host cell

During the HIV budding process, gp120-gp41 complexes are initially expressed at the surface of infected cells and incorporated into the virus envelope to be displayed on its surface as viral spikes. The HIV envelope spike consists of three gp120 and gp41 subunits in combination to form a trimer of heterodimers that mediate binding and entry of the virus into the host cell [147]. Viral entry into a host cell first involves the interaction of gp120 with the CD4 receptor [148]. This induces a conformational change in gp120 through the variable loops so that the co-receptor binding site is formed and exposed [149]. Subsequently the chemokine receptor, CCR5 [150] or CXCR4 [151], is bound and induces a conformational change in the N-terminus of gp41 so that the viral and cellular membranes fuse, enabling the viral capsid to enter the cytoplasm of the cell [149]. The viral core then goes through the process of uncoating and the initiation of reverse transcription of the viral RNA genome.

1.2.3.1.2 Reverse transcription

Once the content of the HIV-1 virion enters the cytoplasm of a newly infected cell, the stage is set for the initiation of reverse transcription. The early steps that lead to reverse transcription are still not completely understood, however transcription occurs shortly after infection because viral DNA can be detected within hours of infection. The main process of converting single stranded RNA (ssRNA) from the virion to linear double stranded DNA (dsDNA) is executed primarily by the enzymatic activities of RT. These are DNA polymerization from the viral genomic RNA and RNase H activity that degrades processed RNA from an RNA-DNA complex. The steps of this process are as follows [60] (Figure 1.5):

- A. Host cell transfer RNA (tRNA), $tRNA_3^{Lys}$, acts as the primer for the synthesis of the minus DNA strand of HIV-1. The 3' end of the $tRNA_3^{Lys}$ is base paired to its complementary sequence in the primer binding site (pbs) in the viral RNA sequence, which is roughly 180 nucleotides from the 5' end of the sequence.
- B. Reverse transcription ensues from the 3'OH group of $tRNA_3^{Lys}$ to produce the “minus strand strong stop DNA” that encompasses the 5' end of the viral RNA, containing the R and U5 regions of the 5'LTR (100 to 150 nucleotides). RNase H degrades the template RNA from the DNA-RNA complex.
- C. The R-region of the “minus strand strong stop DNA” base pairs with its complementary sequence in the R-region of the 3' end of viral RNA (minus-strand transfer), which can occur within the same RNA strand being

processed or at the 3' R-region of the other packaged RNA of the diploid genome. Minus strand DNA continues to be generated up to the 5' pbs while RNase H degrades template viral RNA as DNA synthesis proceeds. However within the RNA sequence of HIV-1 are two polypurine tracts that are resistant to RNase H degradation, one at the 3' end of the RNA (3' ppt) and the other in the middle of the genome, called the central ppt (cppt).

- D. The 3' ppt and cppt that are not degraded by RNase H and function as primers for the initiation of the synthesis of plus strand DNA. Generation of the plus strand DNA starts at the 3' ppt using the minus single-stranded DNA as the template for plus strand DNA synthesis, including the 3' pbs which is complementary to the 3' end of the minus strand DNA.
- E. RNase H removes the tRNA₃^{Lys} to expose the minus strand pbs which binds to its complimentary sequence in the 3' end of the plus strand DNA (plus strand DNA transfer) to form a transient circular DNA and DNA synthesis continues to proceed to the end of both templates, completing the generation of both DNA strands.
- F. Plus-strand DNA synthesis ceases at the end of the minus-strand or when RT reaches the cppt, which results in linear dsDNA with a central plus strand overlap of DNA (~99 nucleotides) known as the central DNA flap. The completed proviral dsDNA contains long terminal repeats at both ends. This flap allows the viral dsDNA to be incorporated into a pre-integration complex (PIC) that facilitates nuclear import of the viral dsDNA. During the integration process, host cell enzymes remove the flaps and repair the gaps in the newly synthesized dsDNA.

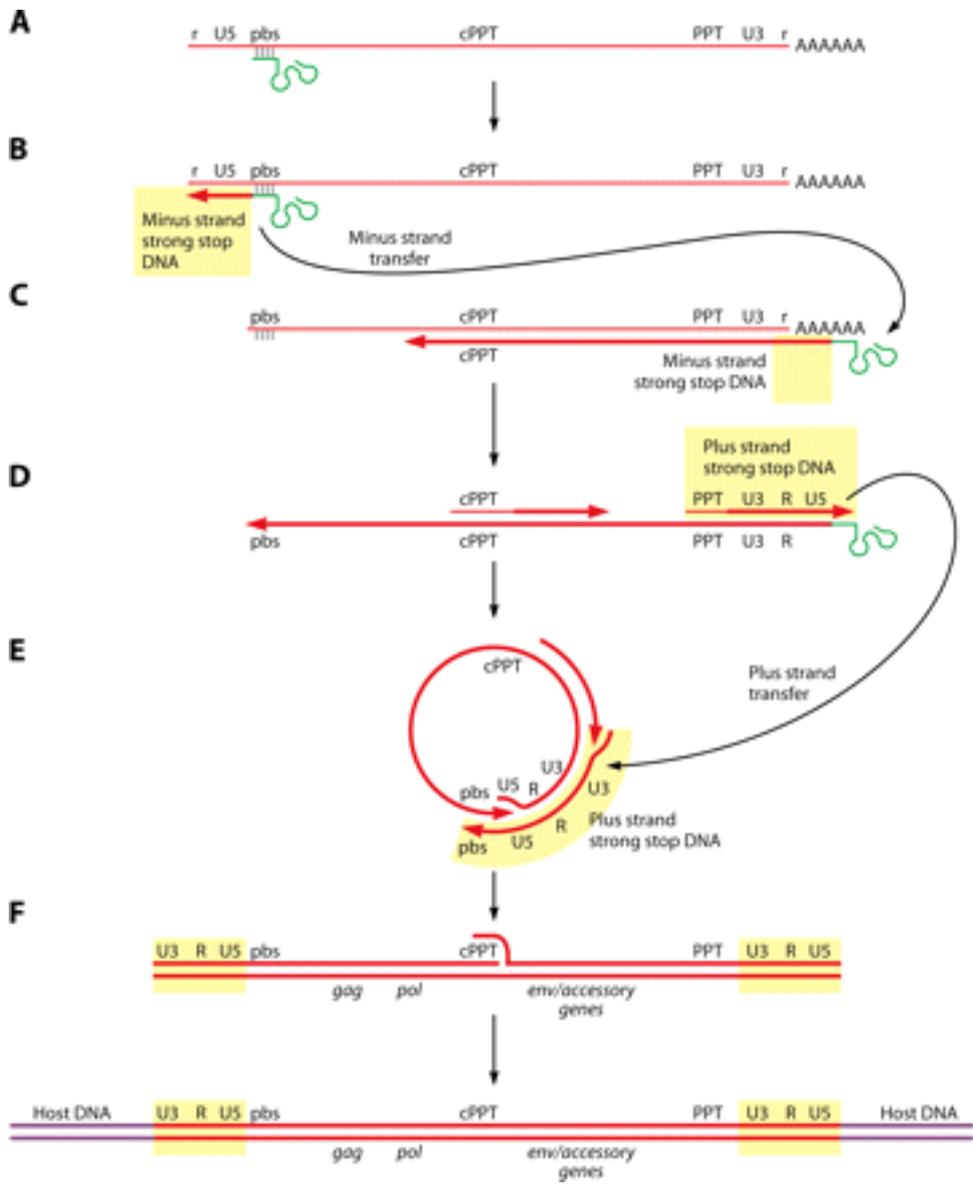


Figure 1.4 The process of reverse transcription of the HIV-1 RNA genome. Figure taken from Coffin et al [60].

1.2.3.1.3 Integration

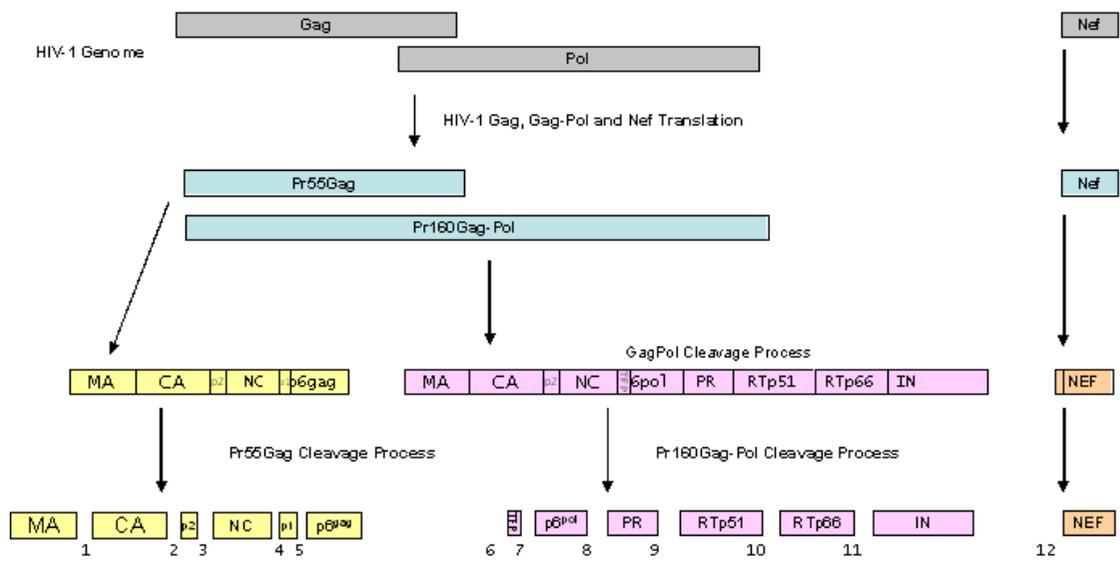
Integration of viral dsDNA occurs in 3 steps [152]: 1) 3' processing, 2) strand transfer and 3) gap repair. 3' processing occurs in the cytoplasm. Integrase cleaves G-T at the 3' end of a conserved CAGT motif on each strand of the viral dsDNA. This results in hydroxyl group overhangs at either ends of the 3' and 5' LTRs. A pre-integration complex (PIC) proposed to contain the processed viral dsDNA, IN, RT, MA and Vpr is imported into the nucleus. A host cell transcriptional co-activator called LEDGF/p75 tethers the PIC to the host cell genomic DNA so that the viral dsDNA is targeted to the host genome [153]. Strand transfer occurs where the two ends of viral DNA are inserted in the host genome at sites that are 5 nucleotides apart. The gaps created in the genome are then repaired by host DNA repair enzymes, removing unpaired viral nucleotides at 5' end of each viral DNA strand, replacing the five bases in gaps created during strand transfer and ligating the viral dsDNA in to the host genome.

1.2.3.2 Late phase

1.2.3.2.1 The action of protease

The Gag and Pol precursor polypeptides, Pr55^{Gag} and Pr160^{Gag-Pol} are the natural substrates of HIV-1 protease [154]. They are cleaved to produce p2, p6 and the structural elements of the virus: MA, CA, NC, the viral enzymes PR, RT, RNase H, IN and a *pol* transframe protein (Figure 1.7). The binding cleft of protease can accommodate 6 to 7 amino acids from its substrate with 3 to 4 amino acids on either side of the scissile bond being cleaved [155].

Gag and Pol proteins are encoded by overlapping reading frames. The *gag* reading frame has its own 'start' (AUG) and 'stop' (TAA) codons and *pol* processing is a result of a -1 frame shift during translation [156]. The ribosomal slippage site where this frameshift occurs is a stem structure at the C-terminus of the sequence that encodes NC, which stalls the ribosome during translation [156]. Protease autocleaves itself from Pr160^{Gag-Pol} at the cellular plasma membrane or just after viral budding [157, 158]. The mechanism by which this process is thought to occur is by dimerization of immature protease within its precursor in *cis*, which creates a partially active catalytic site that cleaves the N-terminus of protease out of its precursor, releasing a fully functional active site with an extended C-terminus [157, 158].



- Pr55^{gag} polyprotein cleavage sites:
1. MA/CA
 2. CA/p2
 3. p2/NC
 4. NC/p1
 5. p1/p6^{gag}

- Pr160^{gag-pol} polyprotein cleavage sites:
6. NC/TFP
 7. TFP/p6^{pol}
 8. p6^{pol}/PR
 9. PR/RTp51
 10. RT/RTp66
 11. RTp66/INT

- Nef protein cleavage sites:
12. Nef

Figure 1.5 Proteins cleaved from the Gag-Pol precursor polypeptide are encoded by HIV-1 *gag*, *pol* and *nef* genes (grey) into their precursor polypeptides (green). The structural proteins (yellow) are formed by cleavage of the Pr55gag polyprotein into matrix (MA, p17), capsid (CA, p24), nucleocapsid (NC, p7), p2, p1, p6gag. The viral enzymes (pink) are formed by cleavage of a second polyprotein, Pr160^{gag-pol}. Although Pr160^{gag-pol} also contains p17, p24 and p2, its C-terminal cleavage products are NC, a transframe protein (TFP), p6^{pol}, protease (PR), reverse transcriptase (RTp51), RT-RNase H (RTp66) and integrase (IN). Nef (orange) is not encoded by *pol* but is also cleaved from its precursor protein by the viral PR (Figure taken from de Oliveira et al [159]).

1.2.3.2.2 Reverse transcriptase

The reverse transcriptase enzyme (RT) is a product of the cleavage of Pr160^{Gag-Pol} by viral protease. Mature HIV-1 RT is a heterodimer with two subunits: p66 (560 amino acids) and p51 (440 amino acids) (Figure 1.4). p51 is a product of protease cleavage of p66. The enzymatic domains of RT, polymerase and RNase H, are found in p66. p51 has identical subunits to p66 at its N-terminus, but does not have the RNase H domain or enzymatic activity; instead, it has a structural role [60].

The crystal structure of RT has been resolved in several states as bound and unbound to its ligand, revealing the right-handed fingers, palm, thumb and connection domains of the polymerase site [160, 161]. The 'palm' of p66 contains the polymerase active site with a nucleic acid binding cleft that exposes three catalytic aspartate residues at amino acid positions 110, 185 and 186. Mutagenesis of any of these residues renders the catalytic site inactive. The suggested mechanism of action of RT is: association of the ssRNA template with the catalytic site of RT. The fingers of the motif closed down around the primer-template and deoxyribonucleotide triphosphates (dNTPs) before the formation of a 3' to 5' phosphodiester bond is formed. Once the bond is formed, the DNA chain lengthens and the finger subdomain relaxes, releasing the pyrophosphate and allowing the next dNTP to be incorporated. The palm and thumb subdomains clamp down on the DNA and the 3'-hydroxyl group (OH) of the primer site in the correct conformation relative to the amino acid residues in the active site of p66 so that the primer-template chain extends for roughly 18 nucleotides between the active sites of the RT and RNase H domains.

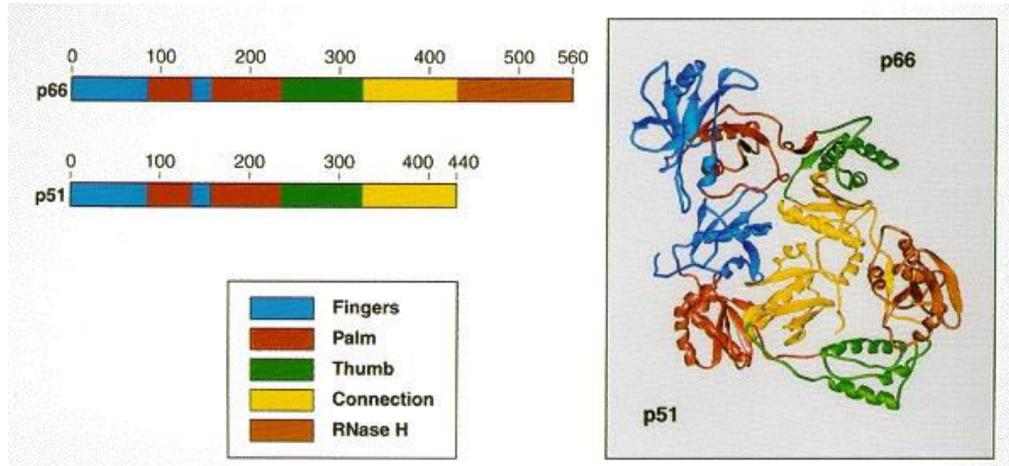


Figure 1.6 The sub-domains of HIV-1 RT. Both subunits, p51 and p66, are shown. The polymerase domains are named based on analogy to a human right hand and color coded: fingers (blue), palm (red), thumb (green), connection (yellow). The RNase H domain in the p66 subunit, is shown in light brown. This figure was taken from Coffin et al [60].

1.2.3.2.3 Integrase

HIV-1 integrase is an enzymatic element of the virion and is responsible for integrating viral DNA into the host genome. It has three domains: (1) an N-terminal domain with a His2Cys2 motif that chelates zinc, (2) a core domain with a catalytic DDE motif which is necessary for enzymatic activity and (3) the C-terminal domain that has an SH3-like fold that non-specifically binds DNA [152].

1.2.3.2.4 Protease

HIV-1 protease is an aspartyl-mediated protease with the active site spread across three amino acids Asp₂₅-Thr₂₆-Gly₂₇ [155, 162-164]. It is 99 amino acids long and contains two identical subunits [163]. A beta-hairpin or “flap” extends from each monomer over the substrate binding cleft [163, 164] and the flexibility of the flap is thought to be essential for enzyme activity by positioning the substrate in the active site [155, 162, 163]. When PIs bind to active site, these flaps protect it from the external environment. Aspartate mediated proteolysis involves the activation of a water molecule between the two Asp₂₅ residues of the homodimer by removing a proton from it so that H₂O⁻ can attack the carbonyl group of the substrate scissile bond which leads to protonation and cleavage of the scissile amide in the substrate [155]. The structure of the viral protease is shown in Figure 1.6.

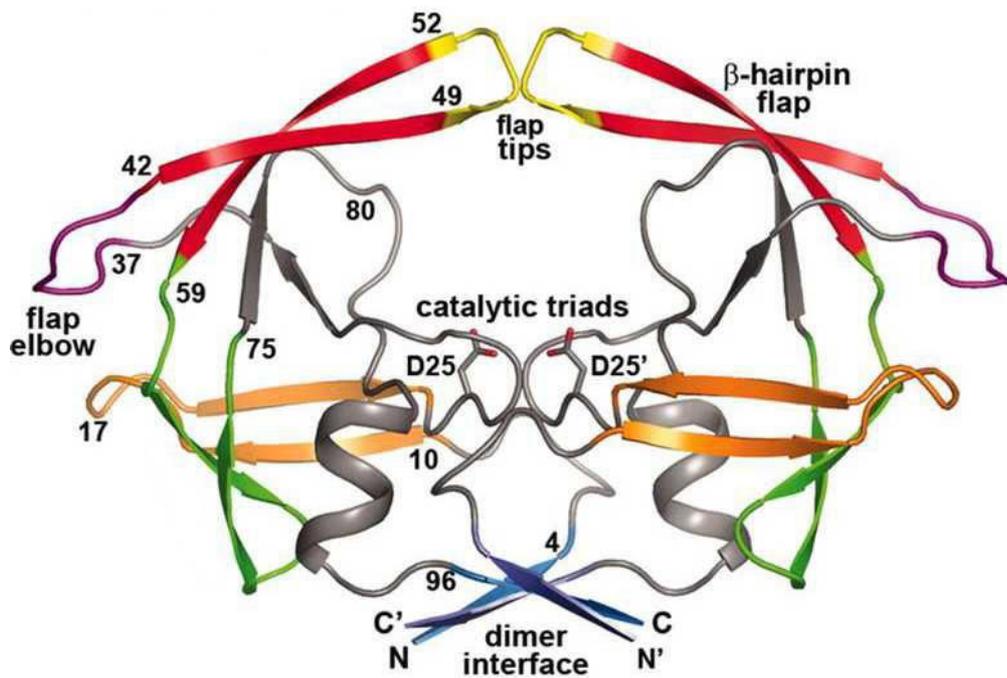


Figure 1.7 Structure of HIV-1 Protease. Colours indicate distinct regions. Red = Flaps: residues 43–58; yellow = flap tips: residues 49–52; pink = flap elbows: residues 37–42; blue = β -strand motif forming the dimer interface: residues 1–4 and 96–99; green = residues 59–75; orange = residues 10–23. Figure adapted from Hornak *et al* [155].

1.3 HIV-1 Antiretrovirals

To date there are 6 HIV-1 antiretroviral drug classes which consist of 24 drugs that have been approved for the treatment of HIV-1 infection by the U.S. Food and Drug Administration (FDA) [165]. Drug classes are based on their point of action/molecular mechanism and resistance profiles. The classes are as follows: (1) nucleoside-analog reverse transcriptase inhibitors (NRTIs), (2) non-nucleoside reverse transcriptase inhibitors (NNRTIs), (3) integrase inhibitors (INIs or InSTIs) (4) protease inhibitors (PIs), (5) fusion inhibitors, and (6) coreceptor antagonists [166]. A schematic of the HIV-1 replication cycle and the point of inhibition/action of each FDA-approved drug/drug class is shown in Figure 1.8.

1.3.1 Targets and mechanisms of action of HIV-1 antiretrovirals

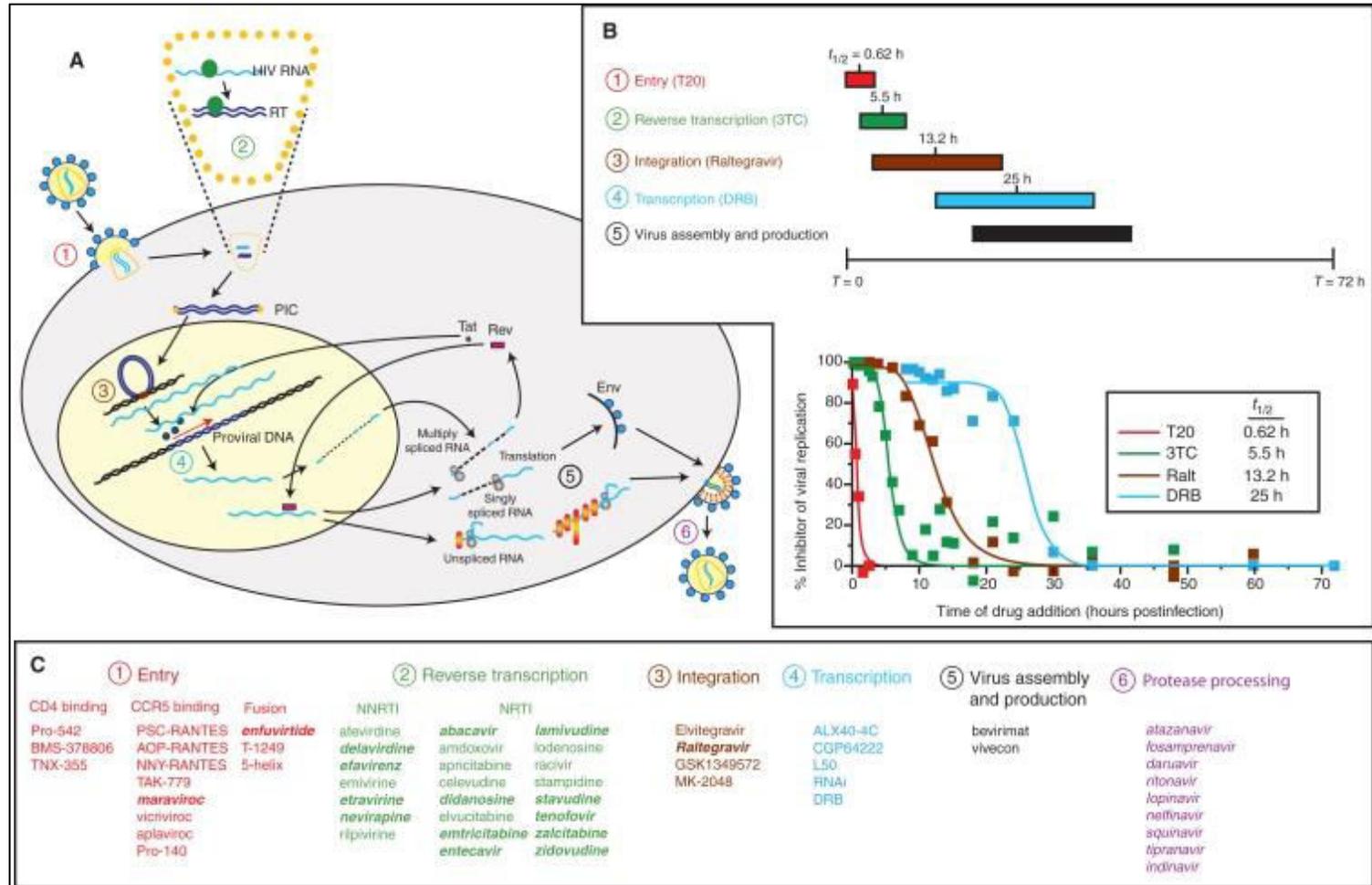
The first step in the HIV-1 replication cycle, (binding, fusion and entry of HIV into the host cell), is the target for antiretroviral agents: attachment inhibitors, chemokine receptor antagonists, and fusion inhibitors. Fusion of the HIV-1 envelope with the host cell membrane is followed by entry of the viral core into the cytoplasm of the host cell and then uncoating of the viral core. Even though the uncoating process is not fully understood, it protects the viral RNA genome whilst permitting access to dNTPs for reverse transcription by the viral RT enzyme and proviral DNA synthesis [167].

RT was the first viral enzyme to be exploited for antiretroviral drug discovery and it is the target of the NRTIs as analogues of native nucleoside substrates. NNRTIs, bind to an allosteric site 10 angstroms (Å) away from the polymerase's active site,

distorting it so that dNTPs cannot easily access the RT active site [160, 168]. There are 12 FDA-approved RTI drugs within these two classes that account for roughly half of all approved ARVs. Although the NRTIs and NNRTIs target different sites and have different mechanisms of action, they both inhibit the same process: DNA polymerization activity of the RT enzyme that inhibits the production of full-length viral DNA. Integrase inhibitors are a relatively newer class of ARV that have been FDA-approved for treatment of HIV infection. They inhibit strand transfer and block integration of the HIV-1 DNA into the cellular DNA.

The PIs block cleavage/proteolysis of the viral polyprotein precursors, a step required for the production of infectious HIV-1 virions. They are the most potent of the approved drug classes and serve as competitive inhibitors to the natural substrate of protease. They also need to be co-administered with a “boosting” agent to inhibit their metabolism through the cytochrome P450-3A4 (CYP3A4) pathway in the liver [169] in order to enhance drug levels. Therefore, PI-based ART includes RTV as a pharmacokinetic enhancer/booster for the active PI in the drug regime [170]. Figure 1.8 shows the drug targets for the ARV classes described above in the context of the life cycle of HIV-1 in a host cell.

Figure 1.8 The HIV-1 life cycle and current or potential targets for antiretroviral drugs (ARV) [166]. (A) Schematic of the HIV-1 life cycle in a susceptible CD4+ cell. (B) Time frame for ARV action during a single-cycle HIV-1 replication assay. (C) Preclinical, abandoned (normal text), or FDA-approved (bold and italic text) inhibitors in relation to specificity of action and drug target. The NNRTI,



Rilpivirine (not in bold) and the integrase inhibitor: Dolutegravir (not included in this figure), are also FDA approved.

1.4 Antiretroviral resistance

1.4.1 Mechanisms of HIV-1 genetic diversity

Genetic diversity of the viral population is key to the selection of drug resistant viral populations in the correct ARV environment. An HIV infected patient generates between 10^9 and 10^{10} virions daily [171, 172] [173]. RNA viral loads of 10^5 to 10^7 copies/ml have been associated with primary infection within one week of patients showing clinical symptoms of infection and represent a mere 10^2 to 10^4 infectious units of HIV [174-177]. The main sources of genetic diversity in an HIV population are (1) the lack of an error-repair mechanism in the viral RT which results in a high mutation rate between 1.4×10^{-5} and 3×10^{-5} nucleotide substitutions per site per generation) and gives rise to every possible point mutation, every day in an infected patient [178, 179], (2) recombination between two RNA strands of the diploid viral genome where each strand encodes different genetic information, and (3) selective pressures induced by antiretrovirals used for treatment or from host immunity.

The main source of mutations in the HIV genome are thought to arise during reverse transcription of viral RNA to DNA by RT because of its lack of exonucleolytic proofreading activity. Other processes in the replication cycle of HIV that can introduce mutations in the viral genome are when integrated proviral DNA is copied by host cell DNA-dependent DNA polymerase when an HIV infected cell replicates and when viral RNA is transcribed from proviral DNA by host cell DNA-dependent RNA polymerase, however their contributions are much less significant because the cellular polymerases have proof-reading activity. Host cellular factors can also influence the generation of mutations, such as APOBEC3G (A3G), which causes G-

to-A mutations in the minus strand of viral DNA via deamination of cytidine [180]. Cellular DNA repair enzymes are also thought to cause mutations in the viral DNA.

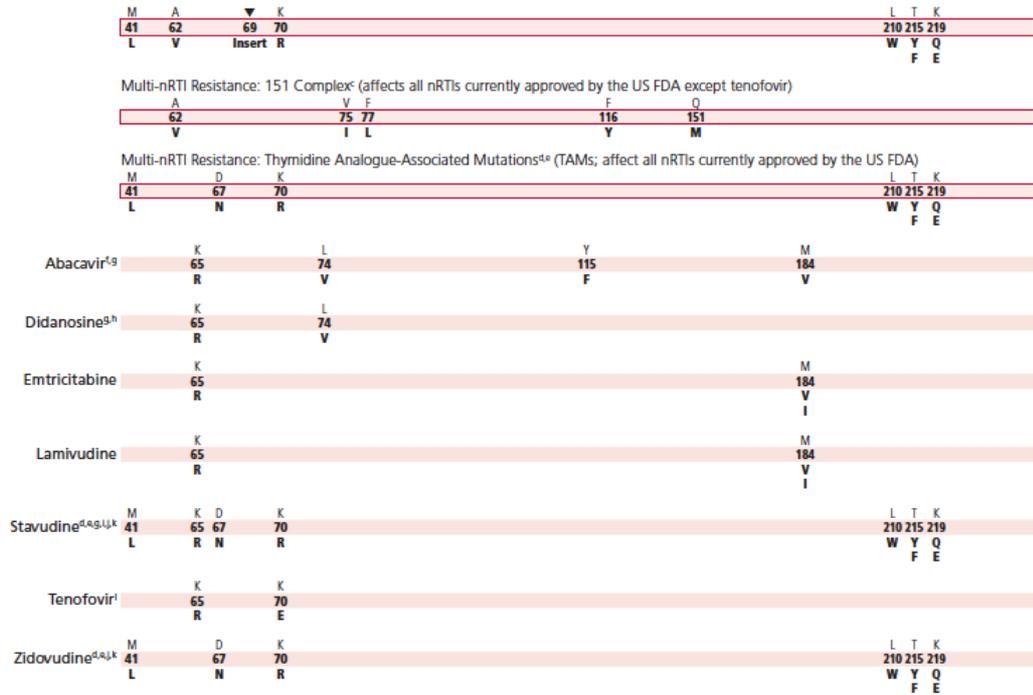
1.4.2 Reverse transcriptase inhibitor resistance

There are two known mechanisms by which NRTI resistance is conferred. The first mechanism is steric hindrance of analogues from being incorporated into the growing DNA chain [168]: the replacement of methionine (M) at position 184 in RT with valine (V) or isoleucine (I), (annotated as M184V/I), is an example of a mutation that causes steric hindrance of NRTIs like 3TC. Position 184 of RT is within a highly conserved motif in the active site of reverse transcriptase [161]. When 3TC tries to be incorporated into the viral DNA chain being generated by an M184V/I mutated RT, the oxathiolane ring of the drug clashes with the β -branch of 184V/I so that it cannot be incorporated into the growing DNA chain [181]. V at this position confers cross-resistance to 3TC, abacavir (ABC) and emtricitabine (FTC) and I at this position confers cross-resistance to 3TC and FTC [182].

The second mechanism by which NRTI resistance is conferred is the removal of thymidine analogues from a terminated DNA chain by pyrophosphorolysis [168]: RT that has thymidine analogue mutations (TAMs) can still allow thymidine analogues like zidovudine (AZT) or d4T (stavudine) to be correctly incorporated into the viral DNA chain. They can be removed by pyrophosphorolysis from the terminated DNA chain [168] and an accumulation of TAMs further enhances this hydrolysis and release of the drug from the terminated DNA chain [183].

NVP and EFV resistance results from mutations in the NNRTI binding pocket of RT, which causes a conformational change that no longer allow the drug to fit in the pocket [160]. RTI resistance mutations in RT are shown in Figure 1.9A and B.

(A)



(B)

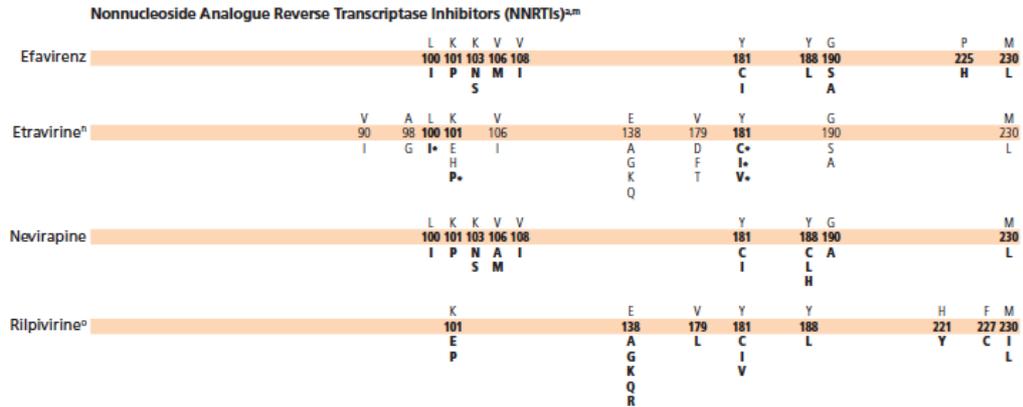


Figure 1.9 Resistance mutations in RT associated with reduced susceptibilities to the RTIs. (A) DRMs that confer resistance to NRTIs and (B) DRMs that confer resistance to NNRTIs. Figures were taken from the International AIDS Society Update on Drug Resistance Mutations 2013 [182].

1.4.3 Protease inhibitor resistance

1.4.3.1 Drug resistance mutations in protease

Mutations in the viral protease that alter the conformation of the active site and its interactions with the enzyme's substrate cause resistance to PIs [168]. PI resistance mutations are categorized as major and minor and the majority are single point mutations [168]. Major and minor mutations have been described in the vicinity of protease's active site as well as outside of the active site. Mutations near the active site are usually major mutations that directly affect the PIs affinity for protease. Mutations outside of the active site are usually minor or silent but influence the stability of the homodimer and the flexibility of the active site.

Major mutations single-handedly cause detectable reductions in susceptibility to one or more PIs. Minor mutations alone usually have a silent phenotype, but in combination with major PI mutations, they enhance PI resistance conferred by the major PI mutations [184, 185] and/or enhance replication capacity of mutated viruses [164, 186]. For example, V82A in PR confers 2- to 4- fold reductions in LPV susceptibility however as minor mutations accumulate in PR, resistance to LPV increases and cross-resistance to other PIs occur [187]. Major and minor PI resistance mutations in PR are shown in Figure 1.10.

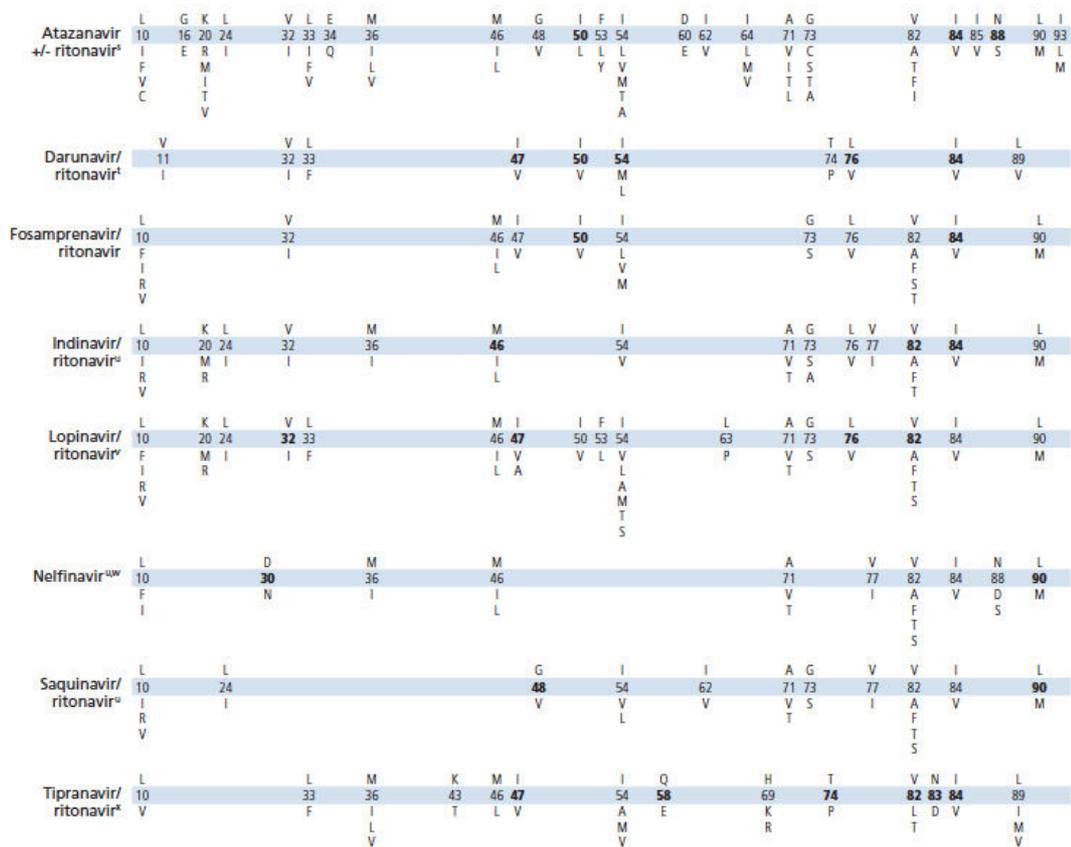


Figure 1.10 Resistance mutations in PR associated with reduced susceptibilities to the PIs. Figure were taken from the International AIDS Society Update on Drug Resistance Mutations 2013 [182]. In bold are major PI resistance mutations which are able to confer reduced susceptibility to PIs one their own. Minor PI mutations are not in bold and are phenotypically silent when present in the absence of a major PI. Minor PI resistance mutations enhance the resistance phenotype of major PI mutations.

1.4.3.2 HIV-1 subtype, antiretroviral resistance and treatment failure

Until 2015, there was no clear demonstration of HIV-1 subtype playing an important role in treatment outcomes and researchers have been divided about its significance in this context: [188-195]. Therefore subtype has not been a consideration for clinical care [103, 191]. In January 2015, Kantor et al [190] were the first to associate HIV-1 subtype to virological failure during ART. They found that patients infected with subtype C viruses “were more likely to fail ART, and to fail earlier than those infected with subtype B viruses”. Virological failure was defined as consecutive viral loads ≥ 1000 copies/ml ≥ 14 weeks after randomization to 1 of 3 treatment arms and viral loads were taken at baseline and then every 8 weeks. The treatment arms were NNRTI or PI based. For the purpose of this project, what is known about subtype C in this context will now be introduced, but it is noted that specific drug resistance mutations have been associated with other HIV-1 subtypes.

The prevalence of NRTI and NNRTI resistance mutations seem to be higher in HIV-1 subtype C [196, 197]. The NNRTI resistance mutation V106M has been associated as most frequently emerged in HIV-1 subtype C and confers cross-resistance to other NNRTIs, compared to V106A [106]. One study found that Etravirine (ETR) preferentially selected E138K as the first mutation in RT for subtypes B, C and CRF02_AG [198]. Another study found that Y181C in RT was the first mutation selected by ETR in subtype C viruses [199]. The TAM K65R has been shown to occur more frequently in subtype C in treatment naïve patients failing TDF-containing ART regimes [200, 201]. Under the selection pressure of PIs, the polymorphism of Methionine at position 89 in PR is thought to preferentially lead to the emergence of threonine at this position [202], which confers high level resistance

to LPV, ATV and NFV [102]. The importance of subtype to mutations in Gag cleavage sites [154] and the viral matrix [154, 203], which are known to be involved in PI susceptibility, are still not well defined. However it is thought that polymorphisms in these regions of Gag may be more frequent in non-B subtypes [191].

1.4.3.3 Contribution of Gag to protease inhibitor resistance

Accessory mutations have been described in the cleavage sites of the viral protease's substrates, the gag and pol precursor polypeptides. The majority of these mutations improve the fitness of mutant virus and a few can confer resistance to PIs [154, 204, 205]. Cleavage site mutations in the gag and pol precursor polypeptides tend to increase the accessibility of the cleavage site for the active site of protease. An exception is the A431V mutation in NC/p1 (of the SP2) cleavage site, which showed to enhance the PI resistance of virus which contain a major mutation in protease [154, 206].

In 1996, Doyon et al [207] analysed the sequence of viruses that were sequentially passaged in the presence of PIs until they displayed up to a 1,500 fold increase in resistance compared to the wildtype. They developed mutations in the CA/p2/NC and NC/p1/p6 cleavage sites of the gag and pol precursor polypeptides after the appearance of major mutations in the viral protease. These cleavage site mutations were Q430R and A431V in the NC/p1 cleavage site and L449F in the p1/p6 cleavage site. When these mutations were reversed by mutagenesis, viral growth experienced a dramatic reduction or was completely abolished, but they did not decrease the ability of the PIs to inhibit the protease's activity. Dam et al [206] went on to show

that A431V augmented PI resistance when associated with V82A; a major PI resistance mutations in protease.

In 2009, Dam et al [206] showed that a patient protease that had V82A in PR in had 21.4-, 16.9- and 15.9-fold increases in resistance to SQV, LPV and ATV. When A431V in the NC/p1 cleavage site was coupled with V82A in protease, the fold change in IC50s increased to >66 for SQV, 56.2 for LPV and 51.1 for ATV.

Verheyen et al [208] also observed compensatory mutations A431V, K436R and I437V of NC/p1 cleavage site as well as L449F/V, P452S and P453L/A in the p1/p6 cleavage site. In 2011, Parry et al [203, 204] identified a triad of mutations (R76K, Y79F and T81A) in matrix that confer intermediate PI resistance (5-10 fold decreased PI susceptibility) in the absence of known resistance conferring mutations in protease.

1.5 Thesis rational

The overall aim of this thesis was to describe the longitudinal development of drug resistance in the viral population of HIV-1 infected children who failed early PI-based ART during the CHER study. In particular, I theorized (1) the development of drug resistance mutations in PR and RT that were linked on the same genome, (2) the persistence of NNRTI resistance during PI-based ART and (3) drug resistant minority species contributing to future ART outcomes by eventually becoming the dominant species under the correct conditions.

Prevention of mother-to-child transmission (PMTCT) has led to significant reductions in vertical transmission [209]. However, children infected with HIV-1 despite this intervention have high levels of non-nucleoside reverse transcriptase inhibitor (NNRTI) resistance [88]. This has led to World Health Organisation recommendations to use protease inhibitor-based (PI-based) triple therapy in such children [45]. Furthermore, the CHER study reported >70% improvement in mortality rates and disease progression in vertically infected children who received early ART [78, 111] compared to those initiating ART according to previous recommendations [44]. Eshleman et al [119] reported that in the absence of ART in children who were HIV infected despite prophylaxis, NNRTI resistance fades from the majority species with time. However, especially in the context of immediate triple therapy, the potential emergence of PI drug resistance on the back of pre-existing NNRTI resistance would signal major implications for longer term antiretroviral efficacy.

The CHER trial was an open-label randomised controlled trial in South Africa [78, 111]. The participants in this trial were randomized to receive deferred or early PI-based ART [40 or 96 weeks of zidovudine (AZT), lamivudine (3TC) and Kaletra® (LPV/r) (Table 3) started within the first 12 weeks of life)] [111]. AZT was replaced with stavudine (d4T) if there were clinical signs of AZT intolerance. Further details of the CHER study can be reviewed in section 1.1.4.1, page 29.

My research cohort was derived from the CHER study; after median follow-up of 4.8 years, 12% (27/230) of the children from the immediate therapy groups had a viral load >1,000 copies/ml [111]. I randomly selected 10/27 children who had a viral load >1000 copies/ml at week 40 of early ART. Table 4 compares the children in my SGA study (VL>1000 at 40wk time point) with the children in the CHER study who had virological failure within a year. One can see that the patients in my study were representative but it was not appropriate to compare the 10 children from the project cohort to the other 17 children because the latter group had virological failures at a median age of 80 months and had been treated for a median of 96 weeks.

Parents provided informed consent for specimens to be collected from the CHER study participants, which were collected and stored as part of the CHER trial. Plasma remaining after viral load testing was used for my analyses and therefore subject to availability.

Study ID	Therapy status			Comments
	Early ART	No ART	Re-started ART	
	Duration (weeks)			
143646	0 – 273			-
141586	0 – 96	97-164		-
130166	0 – 40			d4T, 3TC, LPV/r from Week 40 – 313
131326	0 – 316			-
153716*	0 – 44			-
138506	0 – 40			-
146666*	0 – 99			RTV-superboost from week 96 – 106 of early ART
147636	0 – 96	97 - 164		RTV-superboost from week 48 – 65 of early ART
141806	0 – 96	97 - 164	165-272	-
134102	0 – 96	97 - 164	165 - 298	RTV-superboost from Weeks 88 – 96 of early ART

Table 3. Antiretroviral history for each child participating in this study. Early ART was started within 1 week of birth for each child and consisted of AZT, 3TC and LPV/r. Three children received additional RTV to achieve mg:mg parity with LPV (RTV-superboost) and 1 child had AZT substituted for stavudine (d4T) at week 40 of ART as indicated in the comments section. “*” Indicated those children who died by the end of the CHER trial from non-AIDS related causes.

Variable	Children in CHER with virological failure within a year (n=7)	Children included in single genome study (n=10)
Median age (IQR) at treatment failure in weeks	46.3 (46.1-52.0)	54.6 (54.1-56.6)
CD4%	33.9 (29.5-40.7)	31.8 (30.8-33.7)
CD4 Count	2084 (1455-2829)	2460.5 (1215-2695)
Log ₁₀ HIV RNA	4.6 (4.2-5.9)	5.5 (4.8-5.9)
Median (IQR) length of time on treatment in weeks	41 (40-48)	48 (40-48)
CDC Stage		
CDC Stage B Disease (%)	0 (0)	0 (0)
Severe CDC Stage B Disease (%)	0 (0)	0 (0)

Table 4. Comparison of the children in my SGA study (VL>1000 at 40wk time point) with the children in the CHER study who had virological failure within a year.

This project is the first to use SGS to characterise antiretroviral resistance in children on early PI-based ART as first line therapy. I used SGS to measure the longitudinal diversity PR-RT within the viral populations of each child in my study cohort (n=10). This technique allowed us to identify drug resistance mutation in PR and RT as well as reveal the presence of viral genomes with linked drug resistance mutations. I used phylogenetic techniques to further characterise the evolution of the viral population and I described the drug susceptibility and replication capacities conferred by patient PR-RT with multiple DRMS linked on the same genomes that conferred reduced susceptibility to multiple drug classes. I also explored the population dynamics and timing of compensatory mutations that emerge in *gag* compared to their associated protease inhibitor resistance conferring mutations in PR, to determine if mutations in *gag* were predictive of mutations in PR.

Chapter 2 General Materials and Methods

2.1 Clinical samples

Two anonymised, patient derived RNA samples from confirmed HIV-1 subtype C infected persons provided by the Virology Laboratory, University College London Hospitals NHS Foundation Trust. These were used for the design and optimization of complementary DNA (cDNA) synthesis and all polymerase chain reactions (PCRs). They were called TS4 and TS5 and had previously determined viral loads of 2,400,000 c/ml and 2,253,700 c/ml, respectively.

Plasma samples from ten children failing immediate PI-based triple therapy from the CHER study were used for this project. The sampling time points from these children included at least a baseline sample and a sample taken after 40 weeks of early ART. The sampling time points obtained for each child are listed in Table 5.

Study ID	Plasma sampling time points (weeks of immediate therapy)							
	baseline	12	40	72	96	164	224	298
143646	Y	-	Y	-	-	-	-	-
141586	Y	-	Y	-	-	-	-	-
130166	Y	-	Y	-	-	-	-	-
138506	Y	-	Y	-	-	-	-	-
141806	Y	Y	Y	Y	Y	Y	Y	-
131326	Y	-	Y	-	-	-	-	-
147636	Y	-	Y	-	Y	Y	-	-
134102	Y	-	Y	-	Y	Y	Y	Y
153716	Y	-	Y	-	-	-	-	-
146666	Y	-	Y	-	-	-	-	-

Table 5. Time points from which plasma samples were obtained for ten children failing immediate therapy in the CHER study. “Y” indicates that plasma was obtained and “-” indicates that no plasma was obtained.

2.2 Molecular biology techniques

2.2.1 Viral RNA extraction

HIV-1 viral RNA was extracted from plasma samples using the QIAamp viral RNA mini kit (Qiagen, Valencia, CA): Purification of Viral RNA (Spin Protocol). To inactivate RNAses and lyse virions to release the viral RNA, 140µL of plasma was added to Buffer AVL with 5.6ug of carrier RNA (in AVL buffer) to a total volume of 600ul (carrier RNA protects viral RNA from degradation and increases the yield of RNA by improving binding to the membrane of the spin column). This mixture was incubated at room temperature for 10 minutes. Viral RNA was ethanol-bound to the membrane of the spin column and sequentially washed with Buffers AW1 and AW2. Finally the purified RNA was eluted in 40ul to 80ul of Buffer AVE. The extracted RNA was either used immediately in cDNA synthesis or stored at -80°C for later use.

The concentration range of RNA required for optimum HIV-1 cDNA synthesis followed by single genome amplification was theoretically determined as 20,000 RNA molecules [114]. Therefore 1,000 to 4,000 copies of RNA per microliter (ul) of AVE buffer were eluted into a total volume of 40ul to 80ul. Below is an example of how I calculated the volume of plasma required for RNA extraction in order to achieve 20,000 RNA molecules for cDNA synthesis:

At baseline, Study ID 134102 had a viral load of 535,000 copies / ml

This is equal to 535 copies of RNA / ul of plasma

I used the formula: (Concentration 1) (Volume 1) = (Concentration 2) (Volume 2)

Therefore:

$(535 \text{ copies RNA / ul}) (\text{Volume 1}) = (4000 \text{ copies RNA /ul}) (40 \text{ ul of eluate})$

Therefore Volume 1 = $(4000)(40) / (535) = 299.1\text{ul}$ plasma used for RNA extraction.

Then I used 5 ul of this eluate, which should contain 20,000 RNA molecules, for the cDNA synthesis reaction.

When there was insufficient sample volume coupled with viral loads that were too low to extract the optimal quantity of RNA molecules, the entire sample volume was used to extract the maximum number of RNA molecules in 40ul of eluate. In order to protect plasma samples from contamination with plasmid DNA and the environment

from infectious virus particles, HIV-1 viral RNA extractions were carried out in a Class II Biosafety Cabinet within areas designated for clinical samples only.

2.2.2 Oligonucleotide primer design

Oligonucleotide primers for cDNA syntheses, PCR amplifications and Sanger sequencing were empirically designed by aligning South African HIV-1 subtypes C sequences from the Los Alamos HIV Sequence database in DNA Dynamo Sequence Analysis Software (BLUETRACKER SOFTWARE LTD). With the exception of the pantropic oligonucleotides HIVOut1, HIVOut2, HIVRes1, HIVRes2, which were provided by the Virology Laboratory, University College London Hospitals NHS Foundation Trust and 5'GagOut, which was designed by Parry et al [203]. The empirically designed primer sequence were based on conserved regions upstream and downstream of the genes of interest for PCR amplification. Particular attention was paid to the 3' end of the primers that were designed to anneal to highly conserved regions and preferentially ending in one or more G or C nucleotides with melting temperatures (T_m) between 45°C and 68°C. The T_m for each primer was calculated by the New England Biolabs T_m Calculator, which can be found at <https://www.neb.com/tools-and-resources/interactive-tools/tm-calculator>, because it was the most suitable for Phusion DNA polymerases and primers >13 nucleotides in length. Sequencing primers were designed with the same criteria and 200-300 nucleotides apart in forward and reverse orientations in order to provide overlapping sequencing reads. The details of each oligonucleotide primer are described in their relevant sections.

2.2.3 cDNA synthesis

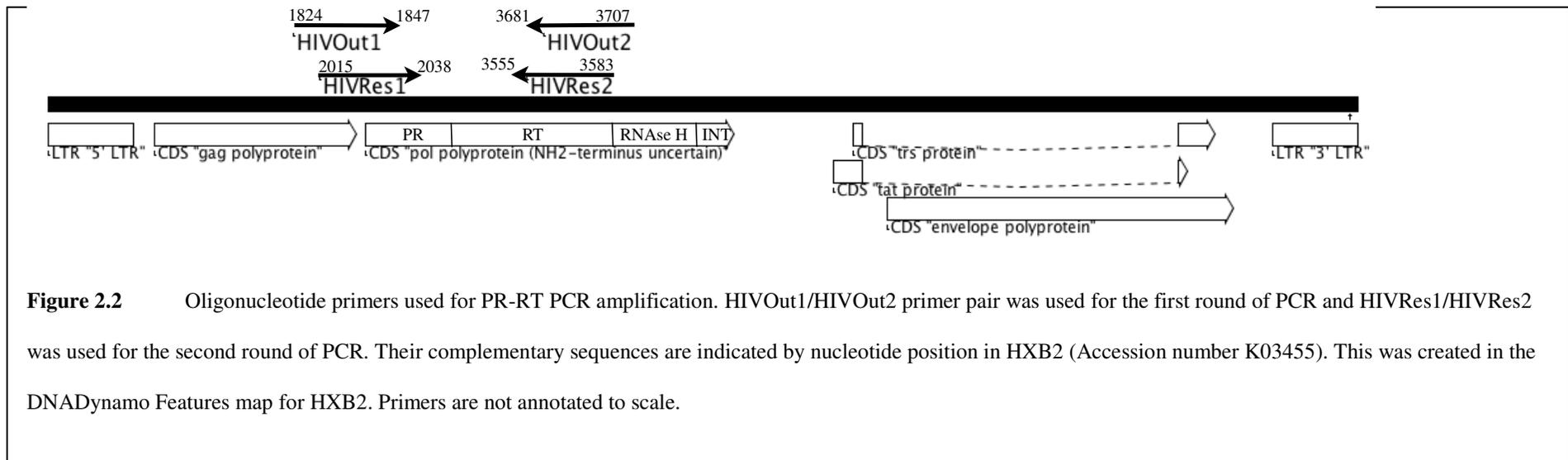
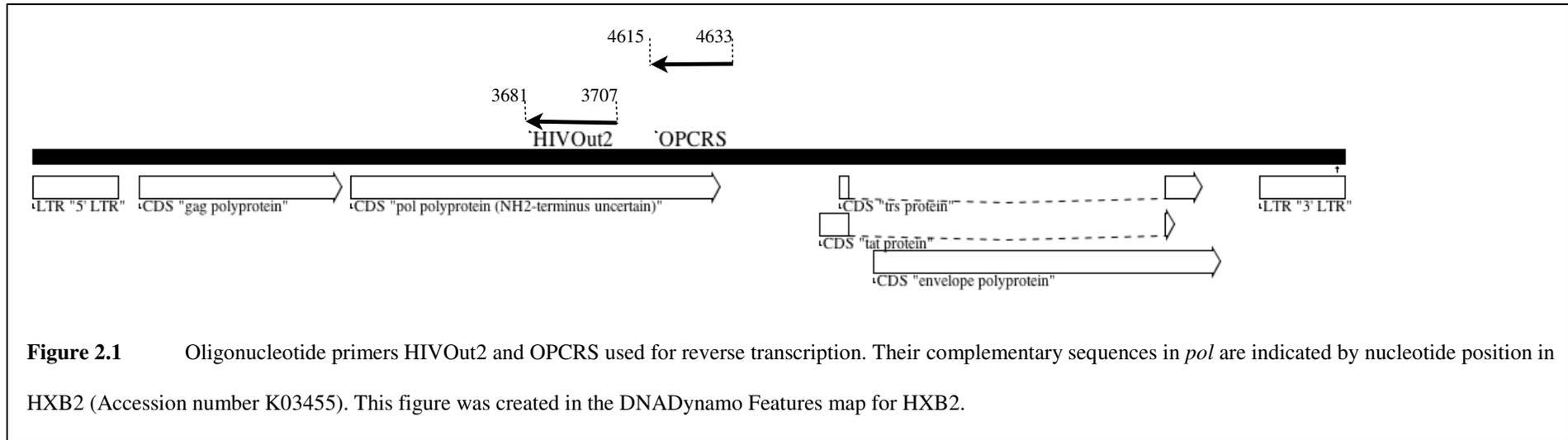
First strand cDNA synthesis from extracted RNA was carried out using SuperScript III (Invitrogen Life Technologies, Carlsbad, CA) and RNase H (2U/μl) treated (Invitrogen Life Technologies, Carlsbad, CA). Briefly, viral RNA was denatured at 65°C for 5 minutes with 0.08μM of a *pol* specific reverse primer and 1mM of dNTP. The denatured and primed RNA sample was immediately incubated at 4°C and mixed with the reverse transcription mastermix containing 0.001M Dithiothreitol, 0.4 unit/ul RNase Out and 1 unit of SuperScript III. cDNA was synthesized at 50°C for 60 minutes and then all enzymes were denatured at 85°C for 5 minutes. Template viral RNA in RNA-cDNA complexes was dissociated by RNase H treatment at 37°C for 20 minutes. Newly synthesized cDNA was either used immediately for PCR amplification reactions or stored at -80°C for later use.

cDNA synthesis of genes encoded by *pol* and *gag* (upstream of *pol*) required a gene specific primer. Two such primers were used: HIVOut2 (5'-AGTCTTTCCCATATTACTATGCTTTC -3') and OPCRS (5'-ATACCTGCCACCAACAGG -3') whose design was based on a consensus sequence of a conserved region of *pol* created by aligning 100 South African HIV-1 subtype C *pol* sequences that were imported from the Los Alamos HIV sequence database: geography search interface [210].

OPCRS was complementary to a sequence at the end of the *pol* gene: nucleotide positions 4615 to 4633 by HXB2 numbering which allowed the 5'LTR, full-length *gag* and *pol* to be reverse transcribed. HIVOut2 was complimentary to a sequence at

the end of the RT subunit in *pol*: nucleotide positions 3681 to 3707 so that the 5'LTR, full-length gag and the RT subunit of *pol* could be reverse transcribed (Figure 2.1).

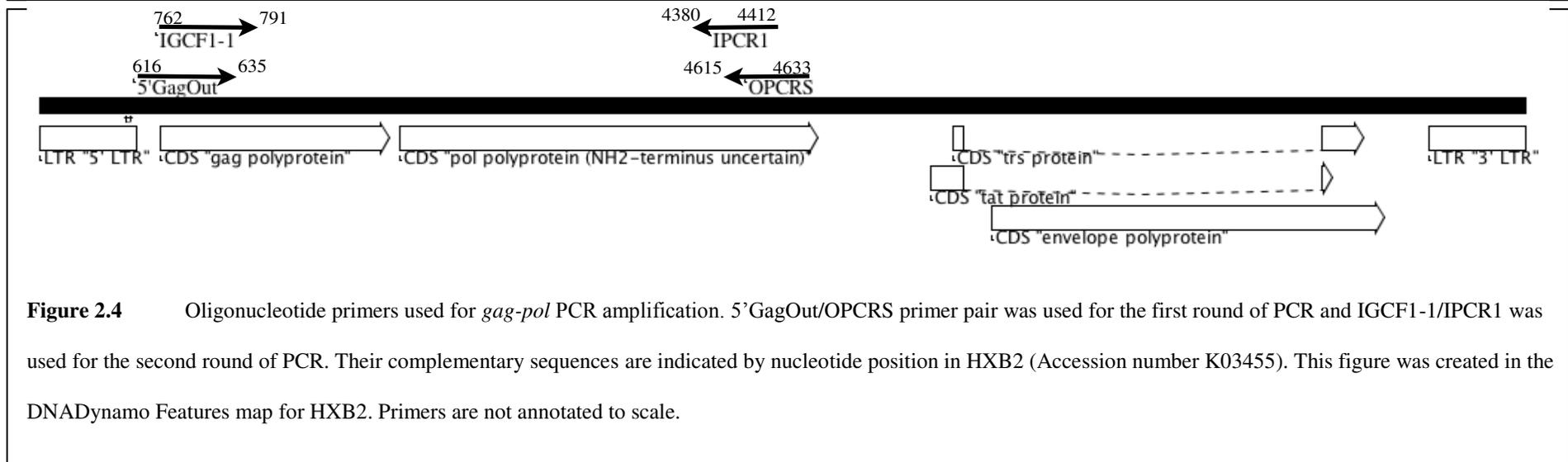
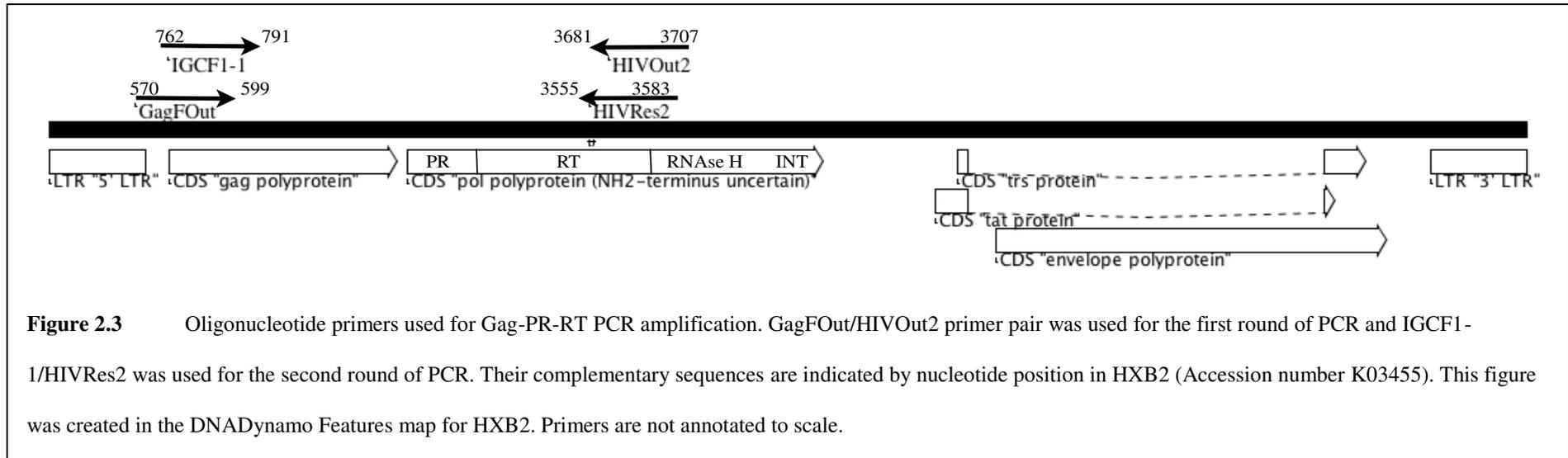
The cDNA synthesis and PCR protocols were validated with test sample TS5. The TS5 RNA inputs used to test the sensitivity of the reverse transcription protocol ranged from 20,000 RNA molecules to 350 RNA molecules. TS5 was used to test the sensitivity of these assays because it was stored at -80°C for 10 years and freeze-thawed twice before use, which was comparable to the age and storage conditions of the project samples and the PCR conditions used were those recommended by the manufacturer for the high fidelity Phusion Flex Hot Start DNA polymerase (New England Biolabs Limited): DNA denaturation at 94 °C, 35 amplification cycles and a final extension at 72 °C for 10 minutes. For bulk PCR amplifications, 1µl of the neat cDNA synthesis reaction (≤ 392 cDNA molecules/ul if 20,000 RNA molecules were reverse transcribed) was subjected to first-round PCR in a final reaction volume of 25µl.



Different concentrations of TS5 RNA were used to validate the sensitivity of the cDNA synthesis protocol using the HIVOut2 primer and the PCR amplification of PR-RT. A nested PCR approach was used with first round PCR primers HIVOut2 and HIVOut1 5'- AATGATGACAGCATGYCAGGGAGT -3' at an annealing temperature (T_a) of 60°C followed by a second round of PCR amplification using HIVRes1 5'- GGAAAAGGGCTGTTGGAAATGTG -3' and HIVRes2 5'- GGCTCTTGATAAATTTGATATGTCCATTG -3' with a T_a of 58°C (Figure 2.2).

PCR amplification of Gag-PR-RT using a nested PCR approach was also done with TS5 using first round PCR primers HIVOut2 and GagFOut 5'- ATTGTGTGACTCTGGTAACTAGAGATCCCT -3' at an annealing temperature (T_a) of 56°C followed by a second round of PCR amplification using IGCF1-1 5'- TTGACTAGCGGCGGCCGCAAGGAGAGAGAT -3' and HIVRes2 with a T_a of 60°C (Figure 2.3).

PCR amplification of *gag-pol* using a nested PCR approach was also done with TS5 using first round PCR primers OPCRS and 5'GagOut 5'- GTGTGGAAAATCTCTAGCAG -3' at an annealing temperature (T_a) of 61°C followed by a second round of PCR amplification using IGCF1-1 5 and IPCR1 5'- CCATARCCCGGGACCACTCTACTTGTCCATG-3' with a T_a of 51°C (Figure 2.4).



All PCR and cDNA synthesis reactions were carried on the G-Storm GS4 Q4 Quad Block Thermal Cycler by Labtech International Ltd. PCR products were visualized by agarose gel electrophoresis that was either manually prepared as 1% agarose supplemented with 0.4mg/ml ethidium Bromide, 1% agarose supplemented with 0.0001% GelRed or using the Invitrogen E-Gel® 96 High-Throughput DNA Electrophoresis System by Life Technologies for products of the predicted sizes: PR-RT \approx 1.7 kilobases (kb); Gag-PR \approx 1.8kb; Gag-PR-RT \approx 2.4 to 2.8 kb; *gag-pol* \approx 3.7 to 4 kb.

To reduce the risks of reaction contaminations, separate plasmid and amplicon-free areas were used to make-up the RT and PCR mastermixes, to add RNA to the RT reactions, to carry out cDNA serial dilutions and to add cDNA to PCR mastermixes. These areas were cleaned with DNase Away (Roche) and UV irradiated for thirty minutes prior to use.

2.2.4 Endpoint dilution PCR

Single genome analysis (SGA) of PR-RT was carried out to determine genetic linkages of drug resistance conferring mutations in full-length protease (PR) and the reverse transcriptase subunit (RT) of the RT/RNase H heterodimer encoded by *pol* as well as the evolution of these mutations in the viral population from baseline. The sampling time points used for PR-RT SGA are those stated in Table 5.

SGA required endpoint dilution of cDNA followed by PCR amplification of PR-RT so that 33% or less of PCR reactions yielded an amplification product. According to the Poisson distribution, the cDNA dilution that yielded PCR products in no more

than 33% of wells contained one amplifiable HIV-1 cDNA template per positive PCR more than 80% of the time [114, 134, 138, 139, 211].

The diluent used for cDNA dilutions was 5mM Tris-HCl. Endpoint dilutions were determined from nine replicate reactions for PR-RT, starting with a 1 in 3 dilution of cDNA which was serially diluted 3-fold to a maximum of 1 in 6561. Particular attention was paid to the distribution of cDNA molecules along the dilution series in order to maintain the reliability of each dilution. This meant that all dilutions in the series were thoroughly mixed to evenly distribute the cDNA. After mixing the cDNA thoroughly, a clean pipette tip was subsequently used to aliquot into the next dilution in the series so that cDNA on the outside of the tip was not transferred into the next dilution in the series. Each aliquot was taken from the surface of the cDNA mixture to avoid transfer of extra cDNA molecules that may adhere to the outside of the pipette tip. A PCR reagent mastermix was made for the number of PCR reactions plus two more reaction volumes. One reaction volume (24ul) was set aside as the negative PCR control and the balance was used for PR-RT amplification from viral cDNA. 10ul of each cDNA dilution in the series was aliquoted into the PCR mastermix, thoroughly mixed and ten 25ul PCR reactions (cDNA + mastermix) were re-aliquoted into ten consecutive wells in a 96-well PCR reaction plate (StarLab Limited, product number E1403-0100) and sealed with an adhesive foil (StarLab Limited) for subsequent PR-RT PCR amplification using the mentioned PCR conditions.

The target endpoint dilution was a yield 22-33% positivity of PR-RT amplification. If this was not achieved from the dilution series, then the Poissonian distribution was applied to the results of the titration to determine the cDNA dilution at which this

positivity was likely to be achieved. Below is a worked example of this calculation for cDNA from the baseline sample of study ID 138506.

Poisson distribution [211]: $F(k; \lambda) = ((\lambda^k) (e^{-\lambda}))/k!$

λ_1 is calculated from a cDNA dilution of 1 in 486 which produced five out of nine (5/9) negative PCR reactions. λ_2 is calculated from the dilution at which to expect 7/9 (88%) negative PCR reactions so that there is greater than an 80% chance that positive reactions are derived from singular HIV-1 genomes. The putative dilution to expect 2/9 (22%) positive reactions (if $-\ln(0.88)$ is used) is calculated with the formula:

$((\lambda_1 / \lambda_2) (\text{fold dilution from which } \lambda_1 \text{ was calculated}))$.

$$\lambda_1 = -\ln(5/9) = 0.587786665$$

$$\lambda_2 = -\ln(0.88) = 0.127833372$$

Therefore:

$$(\lambda_1 / \lambda_2)(486) = (0.587786665/0.127833372) (486) = 2235$$

Therefore cDNA derived from the baseline sample of 138506 should be diluted 2235-fold for a 22% positive yield of PR-RT amplicons.

Once the endpoint dilution was determined, PR-RT PCR mastermix was made for 192 PR-RT reactions (176 PR-RT amplifications from endpoint dilutes cDNA and 16 PCR negative controls) and aliquoted into two 96-well plates for subsequent PR-RT

PCR amplifications and visualization with the Invitrogen E-Gel® 96 High-Throughput DNA Electrophoresis System by Life Technologies.

2.2.5 Sanger sequencing

The nucleotide sequences derived from bulk and single genome amplifications were determined by Sanger sequencing provided by Beckman Coulter Genomics Incorporated. A 25ul PCR reaction yielded ~500ng of non-purified amplicons, therefore a 1 in 5 dilution of non-purified PCR product (~90 to 100ng/ul PR-RT amplicons) and 5uM of each sequencing primer were provided for sequencing. The sequencing primers used for each gene of interest are listed in Table 6.

2.2.6 Sequence analysis and multiple sequence alignments

Sequencing reads (contigs) generated from each sequencing primer were imported into DNA Dynamo Sequence Analysis Software (BLUETRATOR SOFTWARE LTD). The chromatograms for each contig were scrutinized for poor quality data, usually at the 5' and 3' end of each read, which were removed and mixed bases were defined and called as mixtures that occurred in more than one contig and were more than 25% of the dominant peak. For single genome sequence analysis, sequences with mixed bases were disregarded and not used for further analysis because they were not derived from single viral genomes. Any sequences containing premature stop codons were also omitted from the analysis.

Multiple sequence alignments (MSAs) of contiguous nucleotide sequences were also created in DNA Dynamo Sequence Analysis Software with default settings and the

sequences were manually determined to be in the correct reading frame by translating them to amino acid sequences.

Gene	Primer name	Sequence	Nucleotide position	Forward/Reverse
<i>gag</i>	IGCF1-1	TTGACTAGCGGCGGCCGCA AGGAGAGAGAT	762-791	Forward
	S1	CATTATCAGAAGGAGCCAC C	1310-1329	Forward
	S2	TCTATCCCATTCTGCAGC	1414-1431	Reverse
	S3	CACATTTCCAACAGCCCTT TTCC	2015-2038	Reverse
<i>pol</i>	HIVRes1	GGAAAAAGGGCTGTTGGA AATGTG	1895-1918	Forward
	S4	GTAAACAATGGCCATTGA CAGAAGA	2610-2635	Forward
	S5	TCCTAATTGAACYTCCCAR AARTCYTGAGTTC	2797-2828	Reverse
	S6	TGGAAAGGATCACCAGCA ATATTCCA	3006-3031	Forward
	S7	AAGCACTAACAGAAGTAAT ACCACTAACTG	3409-3438	Forward
	HIVRes2	GGCTCTTGATAAATTTGAT ATGTCCATTG	3555-3583	Reverse
	HIVOut2	AGTCTTTCCCATATTA TGCTTTC	3681-3707	Reverse
	S8	GGAAAAGCAGGATATGT TA CTG	3906-3927	Forward
	S9	CCATARCCCGGGACCACAC TCTACTTGTCATG	4380-4412	Reverse

Table 6. Sequencing primers used for Sanger sequencing. The nucleotide position of the complimentary sequence to each primer is indicated as well as the direction of the sequencing read.

2.2.7 Subtype and drug resistance determinations

PR and RT nucleotide sequences were submitted to the Genotyping Resistance Interpretation Algorithm hosted by Stanford University's HIV Drug Resistance Database [102]. PR-RT sequences were also submitted to REGA HIV-1 subtyping tool at <http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool/> for subtype determination.

To identify compensatory mutations in gag that are associated with drug resistance conferring mutations in the viral protease. Table 7 lists these mutations. Patient derived *gag* sequences were translated to their amino acid sequence in DNADynamo and then screened against the mutations in Gag from Table 7.

Unit of Gag	Mutation	
	Associated with PI resistance	Associated with PI exposure
Matrix	76K, 79F, 81A occurring simultaneously	12K, 62R, 75R, 112E
Matrix/Capsid CS	128I/T/A/deletion	132F
Capsid		200I, 219Q/P
SP1		360V, 362I, 363M/F/C/N/Y, 368C/N, 369H, 370A/M/deletion, 371deletion, 373P/Q/T, 374P/S, 375N/S, 376V, 381S
Nucleocapsid		389T, 390A/D, 401T/V, 409K
SP2	431V, 436E/R, 437T/V, 449F/P/V, 452S/K, 453A/L/T	428G, 430R, 451T/G/R
p6		437N, 468K, 474L, 484G, 487S, 497L

Table 7. Mutations in Gag associated with PI resistance and/or PI exposure.

This table was adapted from Fun et al 2012 [154].

2.2.8 DNA quantification

DNA was quantified using the NanoDrop 1000 Spectrophotometer by Thermo Scientific Ltd. DNA was quantified from 1ul volumes of plasmid DNA or purified PCR product DNA according to the manufacturer's instructions. The ratio of absorbance at 260 nm was used to assess the purity of DNA in each sample. A ratio of ~1.8 was accepted as "pure" for DNA.

2.2.9 Addition of restriction endonuclease sites to amplicons by PCR

To facilitate cloning of PCR products into the *gag-pol* expression vectors p8MJ4-HSC, nested PCR primers containing unique restriction enzyme sites were used to introduce these sites at the 3' and 5' ends of the PCR products so that they flank the genes of interest for cloning. The nested PCR primers designed for the addition of these restriction enzyme sites contained the relevant restriction endonuclease site-specific sequence but did not alter the amino acid coding of the fragment. These PCR reactions were also carried out with Phusion Flex Hot Start DNA polymerase (New England Biolabs Limited) and according to the manufacturer's instructions and the same PCR platform outlined in the previous section.

PR-RT amplicons were generated with the first-round PCR primers 5'GagOut and HIVOut2 ($T_a = 56^\circ\text{C}$ for 30 seconds) followed by a second-round of amplification with the primers GagApaF1 and RTRevHpa1 ($T_a = 60^\circ\text{C}$ for 30 seconds). GagApaF1 introduced an Apa I site directly upstream of the translational Start of the PR encoding gene and RTRevHpa1 introduced an Hpa I site flanking the last codon of RT (Figure 2.5).

2.2.10 Cloning into p8MJ4-HSC

PR-RT and *gag-pol* were purified from their PCR reaction mixtures using the Illustra GFX PCR DNA and Gel Band Purification Kit protocol for the isolation and concentration of DNA fragments from PCR mixtures (GE Healthcare Life Sciences) after being generated by PCR amplification so that they were flanked by their respective restriction endonuclease sites as described in the previous section.

2.2.11 Plasmid preparation

2.2.11.1 Apa 1/ Hpa 1 double digestion of p8MJ4-HSC

The *gag-pol* expression vector p8MJ4-HSC (Figure 2.6) was created by Mbisa et al [212] to accommodate RT domain-swapping. It was derived from the *gag-pol* expression plasmid vector p8.9NSX+ [203] with an HIV-1 C-type *gag-pol* sequence that was modified by introducing three restriction enzyme sites including a unique Hpa I flanking RT amino acids 288/289 and included a natural and unique Apa 1 site in the p6 region of *gag* (Figure 2.6).

Wildtype PR-RT was released from p8MJ4-HSC by double digestion with Hpa 1 (New England Biolabs) and Apa 1 (New England Biolabs). Both enzymes were 100% active in CutSmart Buffer (New England Biolabs). One microgram of p8MJ4-HSC was first digested with 2.5 units of Hpa I and 1X CutSmart Buffer (New England Biolabs Limited) in a 10ul total reaction volume for 1.5 hours at 37°C. The singly digested plasmid was purified from the enzymatic reaction using the illustra GFX PCR DNA and Gel Band Purification Kit protocol for the isolation and concentration of DNA fragments from restriction enzyme digests. The purified, singly digested plasmid was eluted in 20ul of Buffer C. A 20ul Apa I digestion mastermix was made

with 20 units of Apa I and 4X CutSmart Buffer (New England Biolabs Limited) was added to the 20ul eluate containing the singly digested plasmid. The new 40ul reaction was mixed gently and thoroughly by pipetting and then divided into two 20ul reactions and Apa I was subsequently allowed to digest the amplicon at 25°C for 30 minutes. The ~1.4kb PR-RT fragment released from the plasmid was distinguished by gel electrophoresis and the vector backbone (~10.7kb) was extracted using the illustra GFX PCR DNA and Gel Band Purification Kit protocol for the isolation and concentration of DNA fragments from DNA-containing agarose gel bands and eluted in 20ul of Buffer C. The Hpa 1/ Apa 1 double digested p8MJ4-HSC backbone was used for cloning patient derived PR-RT sequences.

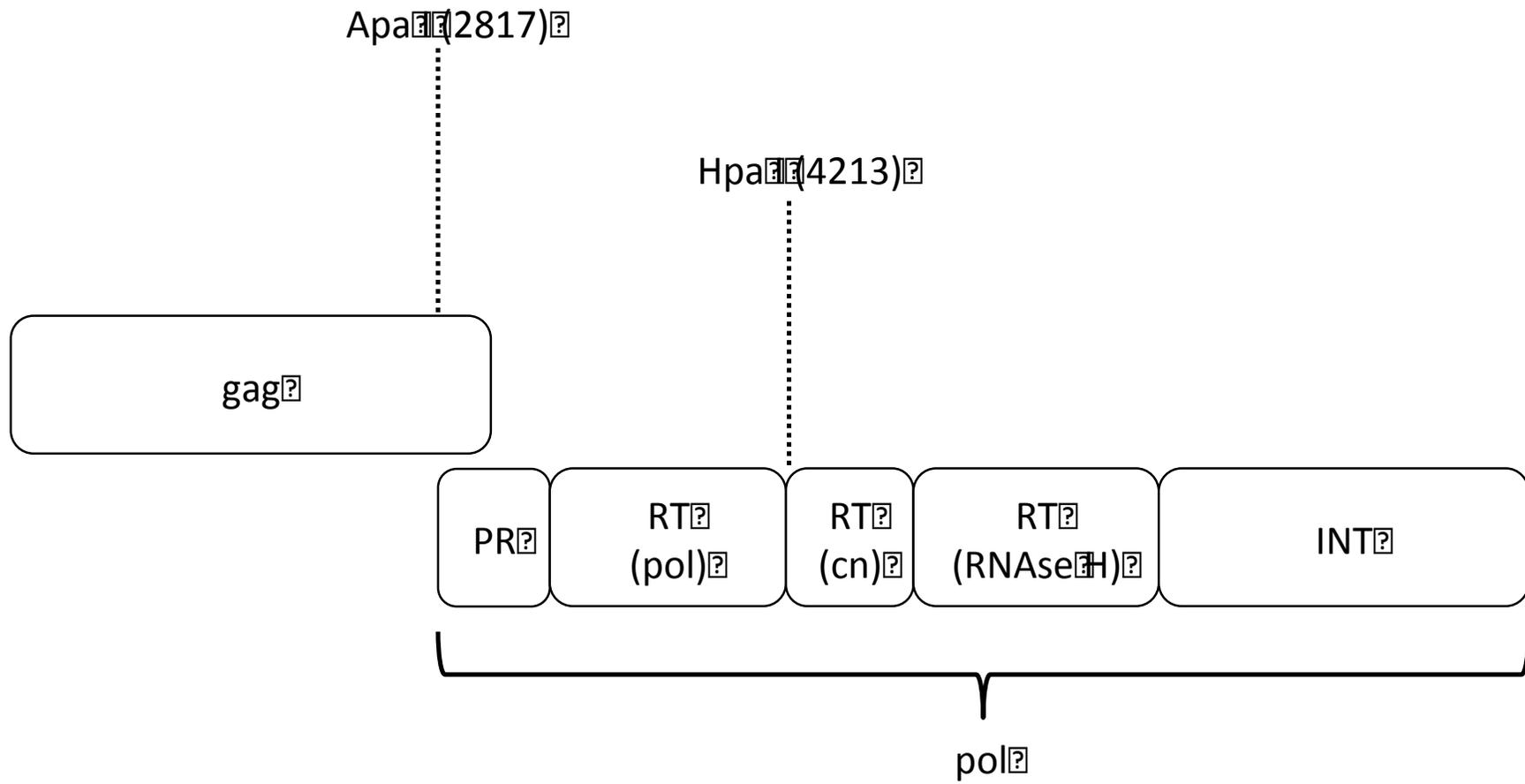


Figure 2.5 The unique restriction sites Apa I and HpaI present in the p8MJ4-HSC gag-pol region that were used for cloning

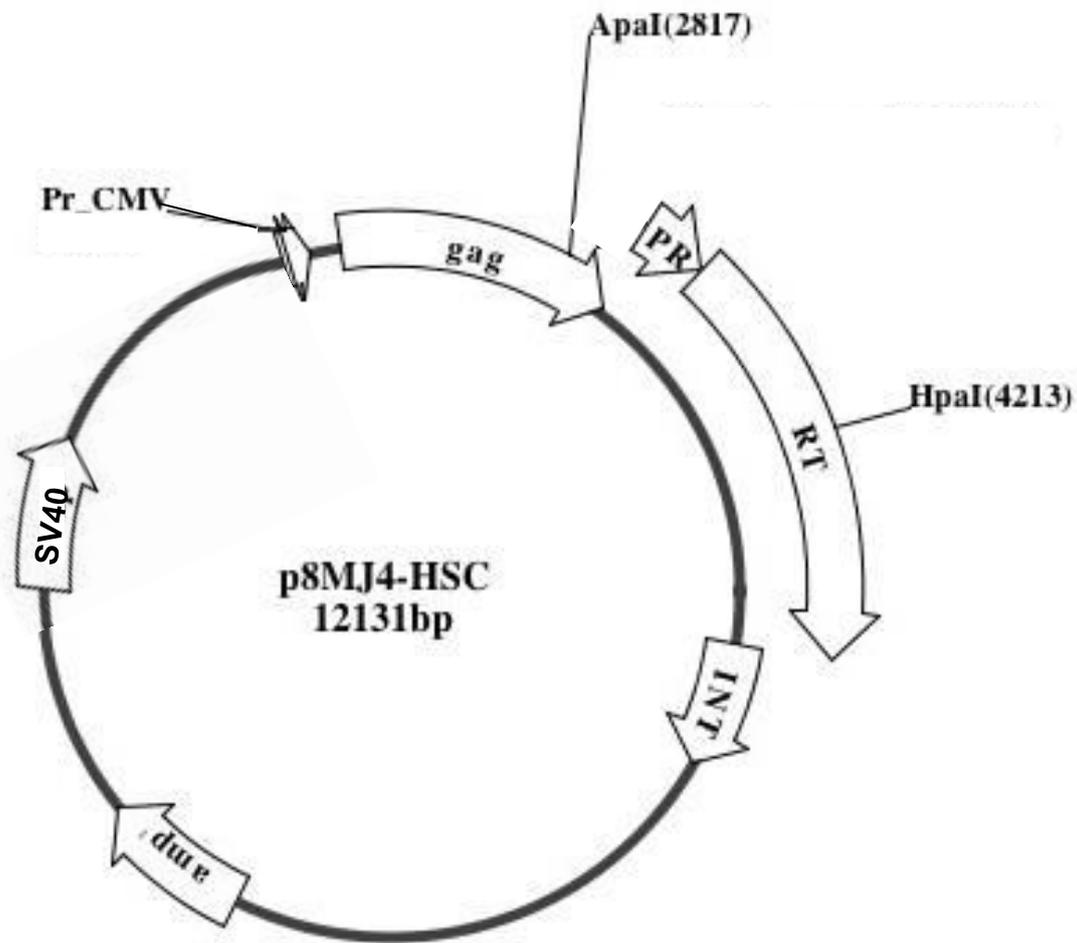


Figure 2.6 Schematic of the p8MJ4-HSC gag-pol expression vector and the unique restriction sites, *ApaI* and *HpaI*, present in its gag-pol region that were used for cloning. This vector expresses the HIV-1 gag-pol gene under the Cytomegalovirus (CMV) promoter. The plasmid vector also contains a Simian Vacuolating Virus 40 (SV40) origin of replication (Ori) which allows replication of the plasmid in mammalian cells that express the large T antigen such as 293T cells. It also contains an ampicillin resistance gene (*amp-r*) for bacterial selection

2.2.11.2 Cloning PR-RT into p8MJ4-HSC

350ng of PR-RT amplicons were digested with 2.5 units of Hpa I and 1X CutSmart Buffer (New England Biolabs Limited) in a 10ul total reaction volume for 1.5 hours at 37°C. The singly digested amplicon was purified from the enzymatic reaction and eluted in 20ul of Buffer C. A 20ul Apa I digestion mastermix was made with 20 units of Apa I and 4X CutSmart Buffer (New England Biolabs Limited) and added to the 20ul eluate containing the singly digested amplicon. The new 40ul reaction was mixed gently and thoroughly by pipetting and then divided into two 20ul reactions and Apa I was subsequently allowed to digest the amplicon at 25°C for 30 minutes and then the double digested PR-RT fragment was purified from the enzymatic reaction.

Apa I/Hpa I double digested PR-RT was ligated into the p8MJ4-HSC double digested backbone (with the same enzymes) in a V:I ratio of 1:7 using 7ng of p8MJ4-HSC vector backbone, 1 unit of T4 DNA ligase and 1X ligation buffer in a 25ul reaction at 16°C for 20 minutes and then 5ul of the ligation reaction was immediately transformed into 50ul of HB101 cells by heat shock at 42°C for 45 seconds and then recovered and cultivated on Luria Broth (LB) agar plates supplemented with 50mg/ml ampicillin (LB amp+) at 30°C. Ten colony-forming units were cultured in LB amp+ broth for 14 hours and plasmids were purified from HB101 cells using the QIAprep Spin Miniprep Kit (250) (QIAGEN®) according to the manufacturer's instructions and purified plasmids were eluted in 50ul of EB buffer.

To determine if PR-RT was successfully ligated into the p8MJ4-HSC backbone, each purified plasmid was double digested with Apa I and Hpa I, as previously described.

If the ligation was successful, a ~1.5kb DNA fragments was released from the recombinant Gag-PR expression vector.

2.3 Phylogenetic and bioinformatics analyses

2.3.1 Phylogenetic reconstruction of PR-RT sequences

The phylogenies of intra-patient viral populations were estimated via a maximum likelihood (ML) approach with the program raxmlGUI version 1.3 [213]. Given a sequence alignment, the ML method determines the probability of observing a tree [214, 215]. The likelihood of all possible trees for a given sequence alignment, i.e the probability of observing a particular tree given the alignment and an explicit model of nucleotide substitution, is calculated and the tree that has the greatest likelihood is selected as the most probable one [216]. The statistical robustness of the trees was evaluated by bootstrap analysis with 1,000 rounds of replication. The phylogenetic trees were visualized and edited using FigTree software version 1.4 provided by <http://tree.bio.ed.ac.uk/software/figtree/>.

2.3.1.1 Inter-patient viral evolution

An MSA was generated for all single genomes derived PR-RT sequences from each sampling time point per patient and was used to reconstruct an ML phylogenetic tree under the GTR model of nucleotide substitution and 1,000 rounds of bootstrapping. The total ML-tree was visualized in FigTree 1.4 and rooted by midpoint rooting to determine if there was contamination between patient data or from external genetic sources such as HIV based plasmids used in the laboratory.

2.3.1.2 Intra-patient viral evolution

For each patient, multiple sequence alignments were generated from single genome derived PR-RT sequences obtained at each available sample time point for that patient and subsequently imported into raxmlGUI version 1.3 to construct ML phylogenetic trees under the GTR model of nucleotide substitution and 1,000 rounds of bootstrapping. Each tree was rooted against an outgroup which was a HIV-1 subtype C PR-RT sequence from the test sample, TS5, used to troubleshoot the PCR and SGS protocols. The criteria I used to choose an appropriate outgroup to predict the direction of evolution within the ML trees were an HIV-subtype C PR-RT sequence from an ART naïve patient. This sequence was related enough to the patient sequences so that it was basal to the rest of the sequences in each tree, but not too closely related that it grouped with test sequences.

2.3.1.3 Mean pairwise genetic distance to measure population diversity

I used the Molecular Evolutionary Genetic Analysis (MEGA) software version 5.2 [217] to calculate the mean pairwise genetic distances (MPDs) of PR-RT sequences within and between sampling time points in each child. In this case, I used the Tamura and Nei 1993 nucleotide substitution model, which was determined as the best fit model for this data by MEGA (lowest Bayesian Information Criterion, therefore highest posterior probability). Regression analysis was used to determine if the number of sequences obtained at each time point affected the estimation of genetic distances. I determined whether differences in the mean number of nucleotide substitutions per site in PR-RT MPDs between consecutive time points were significant using an unpaired two-tailed Student's t-test. MPD was expressed as number of nucleotide substitutions per site of all PR-RT sequences.

2.3.2 Assessment of recombination in PR-RT

Evidence of recombination between PR-RT sequences derived from single genomes over time was determined using the Single Breakpoint Analysis (SBP) and Genetic Algorithm Recombination Detection (GARD) from the online Datamonkey software package at <http://www.datamonkey.org>. Significant breakpoints were reported for P values <0.05 .

2.3.3 Intra-patient analysis of selection pressure on the HIV-1 PR and RT genes

Intra-patient selective pressures on HIV-1 PR-RT and gag-PR-RT were determined with the Datamonkey software package. For positive selection analyses, the rate of non-synonymous substitutions per non-synonymous site (dN) over the rate of synonymous substitutions per synonymous site (dS), dN/dS or ω , was calculated using three different algorithms. If dN/dS = 1 then this suggested neutral selection, if dN/dS is <1 this suggested negative selection because there were more synonymous substitutions than non-synonymous ones (indicating that non-synonymous changes at that site are removed from the population) and if dN/dS is >1 , positive selection is suspected because there were more non-synonymous substitutions than synonymous ones (Vandamme et al., 2009). The three different algorithms used to determine ω within each patient viral population were FEL (Fixed Effects Likelihood), SLAC (Single Likelihood Ancestor Counting) and FUBAR (Fast Unbiased Bayesian AppRoximation).

All three methods calculate dN/dS ratios for each codon in a given sequence alignment: FEL, which directly estimates nonsynonymous and synonymous

substitution rates at each site [218]; SLAC, which estimates the number of non-synonymous and synonymous substitutions that occurred at each codon in an alignment, by reconstructing the most likely ancestral sequences and counting substitutions using a weighting scheme [219]; FUBAR, which detects selection under a model which allows substitution rate variations from site to site and calculates the mean posterior distribution of synonymous (α) and non-synonymous (β) substitution rates [220].

Multiple sequence alignments of PR-RT accompanied by corresponding ML-trees generated from raxmlGUI version 1.3 were uploaded to the online Datamonkey platform. Substitution models were determined using the automatic substitution model selection tool, which selected the HKY85 model for PR-RT sequence alignments. SLAC, FEL, and FUBAR algorithms were then run with the selected substitution model with confidence intervals of 1.0 and a significance level of <0.05 (SLAC and FEL) or a posterior probability of 0.95 (FUBAR).

2.3.4 Co-evolution analysis

I also used the algorithm Spidermonkey [221] at <http://www.datamonkey.org> to determine if positively selected sites by SLAC, FEL or FUBAR co-evolved, i.e. if the evolution of amino acids at any pair of positively selected were dependent on each other during the course of protein evolution

2.3.5 Position Specific Scoring Matrix

I used the perl-based script “aa_freq.pl” developed by Professor Simon Watson (unpublished) to produced a position specific scoring matrix (PSSM) from an MSA of

amino acids from the 891 HIV-1 Subtype C Gag sequences from treatment naïve children from Sub-Saharan African children that were available in the HIV Los Alamos sequence database on June 1st 2014. I used the PSSM to determine the natural variation of amino acids found at these key positions in Gag for patients from my study cohort using population sequence analysis.

2.4 Tissue culture

2.4.1 Pseudovirus production

Pseudoviruses were generated from gag-pol expression vectors with successfully cloned patient PR-RT fragments (see section 2.2.10) from patient derived virions using the method developed by Parry et al in 2009 [203]. In brief, Human Embryonic Kidney 293T cells (293T cells) were co-transfected using FuGENE® 6 Transfection Reagent (Promega UK) with the gag-pol expression vectors p8MJ4-HSC, which expresses patient derived gag-pol fragments with the pCSFLW and pMDG vectors. The pseudoviruses produced contain the HIV-1 *gag* encoded structural elements, the HIV-1 *pol* encoded enzymatic elements surrounded by a VSV-G envelope from pMDG and a genome encoding the luciferase reporter gene from pCSFLW.

2.4.2 FuGENE® 6 transfection for pseudovirus production

Four millilitres Gibco® Dulbecco's Modified Eagle Medium supplemented with 10% foetal calf serum (DMEM) (Life Technologies Ltd) containing 5×10^5 293T cells were seeded into each well of a 6-well plates the day before transfection. The next day, 6ul of FuGENE® 6 was added drop-wise into 70 μ L Opti-MEM® I Reduced Serum Media (Opti-MEM®) (Life Technologies) in a 1.5ml eppendorf which was mixed by flicking for each well. The mixture was left for 3-5 min and then a 10 μ L

mixture containing 300ug of the gag-pol expression vector, 300ug of pMDG and 500ug of pCSFLW was added drop-wise into the Opti-MEM®/ FuGENE® 6 mixture and mixed by flicking and left for 15 minutes, in which time the 2ml of DMEM in each well of seeded 293T cells was refreshed with 4ml of DMEM. The Opti-MEM®/ FuGENE® 6/DNA mixture was then added drop-wise on to the 293T cells, mixed by swirling and then incubated at 37°C with 10% CO₂ for 48 hours. Pseudovirions were subsequently harvested from the supernatant using 0.45µm filters and either used immediately for further experiments or stored in at least two aliquots at -80°C.

2.4.3 Single-replication cycle drug susceptibility assay

HEK 293T cells were infected with the pseudovirions produced from the triple-plasmid co-transfection of producer 293T cells as described in section 2.4.1 and luciferase mRNA is integrated into the host cell genome so that the expression of luciferase can be measured as an indicator of infection from transduced cells. This process is exploited for drug susceptibility assays by producing pseudovirions in the presence of antiretrovirals and the measure of luciferase expression served as an indicator of resistance to the drug where the concentration of the drug that induced 50% infectivity compared to a baseline (IC₅₀) can be calculated.

This is a single-replication cycle assay that can be carried out in the Category 2 laboratory because harvested pseudovirions do not harbor gag-pol mRNA after transfection, therefore no further virions were generated.

2.4.4 Replication capacity

The replication capacity (RC) of pseudoviruses produced by each chimeric vector was determined from titrations of serially diluted viruses on 293T cells for 48 hours. Infection was determined by measuring the luciferase expression of infected target cells. An enzyme linked immunosorbent assay (ELISA) (Reverse Transcriptase Assay, colorimetric®Roche; Catalogue number 11468120910) was used to measure the reverse transcriptase activity of each pseudovirus. RC was measured as RLU/ng of RT activity and expressed as % of the WT for each virus.

2.4.5 Protease inhibitor susceptibility assay

The protocol for pseudovirus production in the presence of serially diluted PIs is outlined in Figure 2.7. As indicated in Steps 2 to 3 of Figure 2.7, producer cells generated pseudovirus in the presence of PIs, where PIs were serially diluted horizontally across the 96-well plate with the highest concentration of antiretroviral (top drug) starting at the second well of the plate (A2-H2) (Step 2). The first well in the plate (A1-H1) was always designated as the 293T cell-free control and the final well in the series was always designated as the virus free control (A12-H12). One hundred microliters of producer cells that were transfected the day before were added to each well of the plate (Step 3). The pseudovirus produced in the presence of each concentration of drug were used to infect fresh 293T cells (steps 4 to 5) and luciferase activity in transduced cells was measured as relative light units (RLU) (step 6) (*Steady-Glo*® Luciferase Assay System by Promega). The HIV-1 C-type *gag-pol* expression vector p8MJ4-HSC was used as the baseline RLU from which the IC₅₀ for each patient derived pseudovirus was calculated. The PIs used in this project were

LPV, SQV, NFV and DRV. The top drugs used for each antiretroviral are indicated in Table 8.

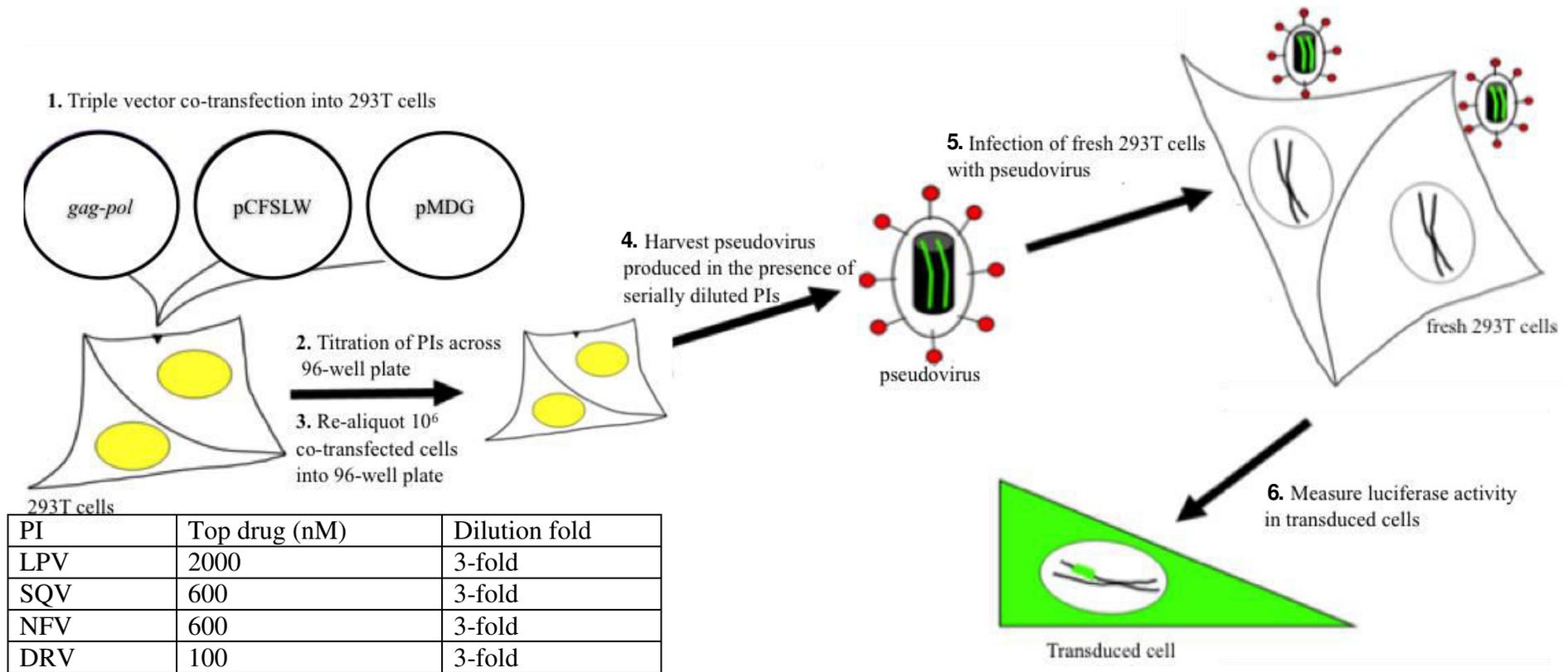


Table 8. PIs concentrations used in drug susceptibility assays.

Figure 2.7 PI susceptibility assay. (1) 293T cells were triple transfected as described in section 2.4.1 and 2.4.2. (2) PIs were titrated across a 96-well plate as described in Table 8 and section 2.4.5 (3) Transfected cells were re-aliquoted into the serially diluted PI from step 2. (4) Pseudovirus produced in the presence of PIs were harvested (5) and used to infect fresh 293T cells. (6) SteadyGlo (Promega) was added to each well to lyse transduced cells and measure their luciferase activity as RLU.

2.4.6 Reverse transcriptase inhibitor susceptibility assay

The protocol for pseudovirus production in the presence of serially diluted RTIs is outlined in Figure 2.8. Prior to starting the RTI susceptibility assays, the harvested pseudovirus from step 2 of the protocol was used for a retroviral titration in the absence of antiretrovirals to determine the concentration at which 10^6 RLU were generated for each test sample (step 3). Harvested pseudovirus from a 2nd aliquot stored at -80°C was diluted to the concentration that yielded 10^6 RLU in the retroviral titration and used to continue with the RTI susceptibility assay (step 4). RTIs were serially diluted horizontally across the 96-well plate and included cell-free and virus-free controls in the same plate format on the plate described for the PI susceptibility assay (step 4) and luciferase activity in transduced cells was measured as relative light units (RLU) (step 5). The HIV-1 C-type *gag-pol* expression vector p8MJ4-HSC was used as the baseline RLU from which the IC₅₀ for each patient derived pseudovirus was calculated. The details of the top drug and fold-dilutions for the antiretrovirals used in this project are described in Table 9 for the NNRTIs: NVP and EFV and in Table 10 for the NRTIs: 3TC, AZT, ABC and ddI.

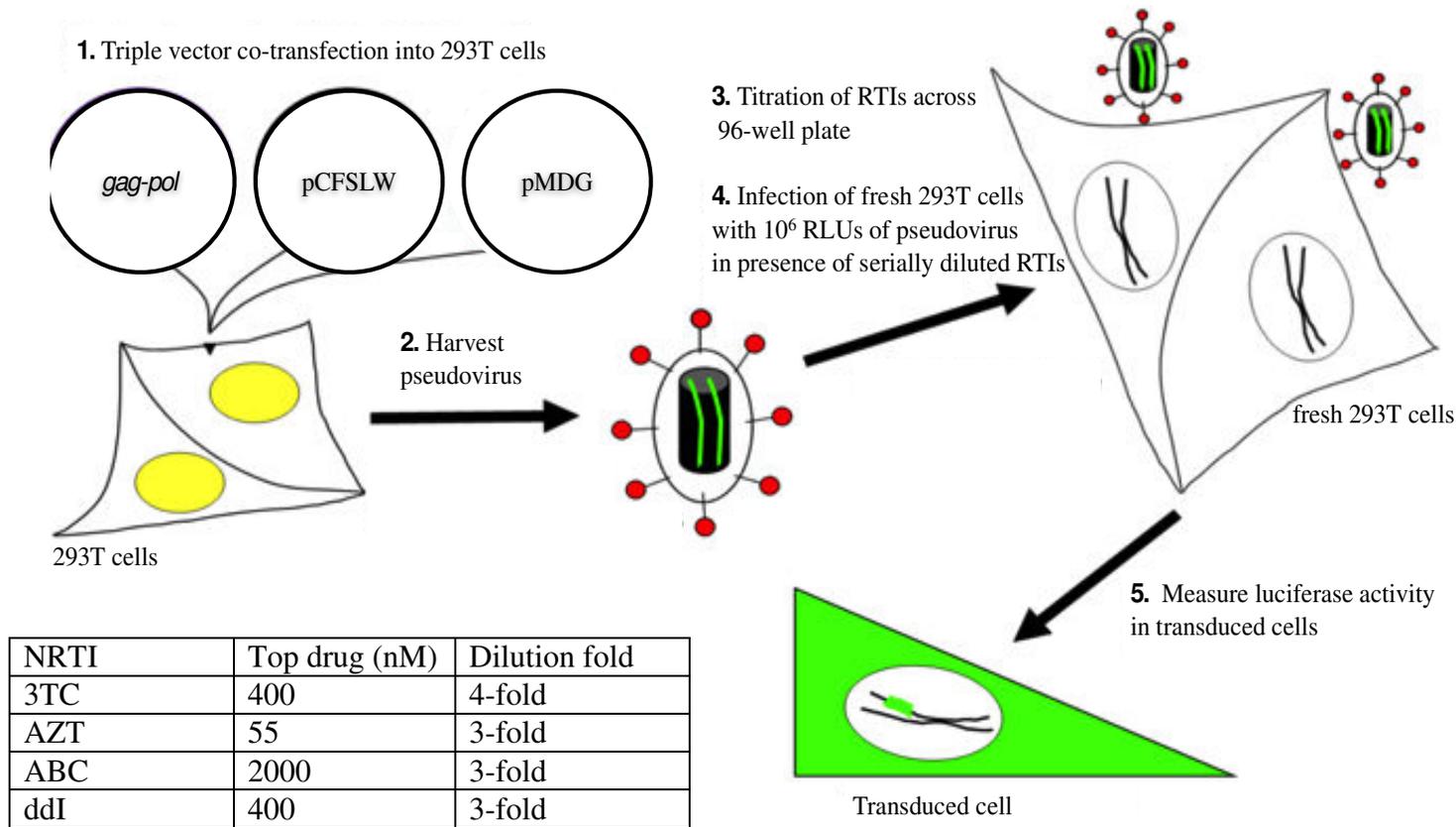


Figure 2.8 RTI susceptibility assay. (1) 293T cells were triple transfected as described in section 2.4.1 and 2.4.2. (2) Pseudovirus from transfected cells were harvested. (3) RTIs were titrated across a 96-well plate as described in Tables 9 & 10 and section 2.4.6 (4) Fresh 293 T-cells were added to the RTI from step 3 and infected with harvested pseudovirus from step 2. (5) SteadyGlo (Promega) was added to each well to lyse transduced cells and measure their luciferase activity as RLU.

Table 9. Details of NRTIs used in drug susceptibility assays.

NNRTI	Top drug (nM)	Dilution fold
NVP	300	3-fold
EFV	100	3-fold

Table 10. Details of NNRTIs used in drug susceptibility assays.

Chapter 3 Results

3.1 Single genome analysis reveals linked, multi-class drug resistance in HIV-1 infected children from the Children with HIV Early Antiretroviral (CHER) study

In this chapter I characterised drug resistance in ten children from the CHER study who acquired HIV infection despite NVP prophylaxis and subsequently failed immediate PI-based therapy. I described the trends in viral load for each child in the context of ART status over time. I then compared bulk sequence analysis to SGS results and used the latter technique to reveal the presence of viral genomes with genetically linked drug resistance mutations that confer reduced susceptibility to multiple drug classes.

After median follow-up of 4.8 years, 12% (27/230) from the immediate therapy groups had a viral load >1,000 copies/ml [111]. Children experiencing treatment failure at the first time point (40 weeks) were first identified. Then those with sufficient samples at subsequent time points of interest were selected and from these, ten were randomly selected. Plasma remaining after viral load testing was used for my analyses and therefore were subject to availability. Plasma was available for 10/10 children before ART was started and at week 40 of ART. 2/10 children who were viraemic at week 72 of ART had plasma samples available at this time point and 3/10 children had plasma samples available at week 96 of ART. I obtained plasma samples after week 96 of ART from 2 children who developed multiple linked multi-class drug resistance in the viral population by week 96 of ART (141806 and 134102) and

from another child (147636) for whom I did not detect such drug resistant variants by week 96 of ART.

I validated the sensitivity and specificity of reverse transcription, cDNA dilution and PCR protocols required for bulk sequence analysis and SGS of PR-RT. I reconstructed the phylogeny of all bulk and SGS sequences generated for each child from all sampling time points available for each child to determine if sequences from each child clustered together. If sequences clustered together by the child from which they were derived, then I used this result as confirmation that there was no external genomic contamination or cross-contamination between patient samples. The PR-RT consensus sequences were submitted to the HIV Drug Resistance Database [102] for HIV-1 subtyping and DRM identification. I used a Fisher exact test ($p < 0.05$ for significance) to identify those codons for drug resistance mutations whose change in frequency between two sequential time points was significant [222].

3.1.1 Results

3.1.1.1 Participants

8/10 children received perinatal NVP for PMTCT. One child did not receive this prophylaxis (Study ID 134102) while this information was not recorded for a second child (Study ID 147636). All mothers received a single dose of NVP at the onset of labour.

Continuous ART was given to 3/10 children (Study IDs 143646, 130166 and 131326) until the end of the CHER trial and their viral loads were suppressed to undetectable

levels (<400 copies/ml) at week 96 of early ART and continued to be suppressed until the end of the trial. Early ART was given to 3/10 children (Study IDs 153716, 138506 and 146666) for 0-40 weeks, 0-44 weeks and 0-99 weeks respectively. These children's viral load histories did not pass the last recorded viral load time point and they were also viraemic at the last recorded time point because Study IDs 153716 and 146666 died of non-AIDS related causes and Study ID 138506 defaulted from the clinic. 4/10 children (Study IDs 141586, 147636, 141806 and 134102) received early ART for 0-96 weeks and only 1 child (Study ID 147636) had an undetectable viral load at week 96. At the end of an ART free period, (weeks 97-164), the viral load reached undetectable levels in only 1 child (Study ID 134102). ART was restarted in 2 of these children, (Study IDs 141806 and 134102), from weeks 165-272 in Study IDs 141806 and weeks 165-298 in Study ID 134102. The viral load of Study ID 141806 became undetectable during re-started ART at week 265 and remained suppressed until the end of the CHER trial, while Study ID 134102 remained viraemic during re-started ART. The viral load histories of each child are shown in Table 10.

Study ID	Weeks										Therapy status		
	0	24	40	72	96	164	224	265	281	298	Early ART	No ART	Re-started ART
	Viral Load (copies RNA/ml blood)										Duration (weeks)		
143646	<u>>750,000</u>	<50	<u>58,200</u>		<400	<400	<400				0-273		
141586	<u>>750,000</u>	<400	<u>729,000</u>	<400	11,040	6,050					0-96	97-164	
130166	<u>>750,000</u>	<400	<u>4,890</u>	640,000		<400	<400			<400	0-313		
131326	<u>318,000</u>	44,000	<u>295,000</u>	<u>56,600</u>	<400						0-316		
153716*	<u>>750,000</u>	<400	<u>4,100</u>								0-44		
138506	<u>617,000</u>	>750,000	<u>>750,000</u>								0-40		
146666*	<u>>750,000</u>	5,400	<u>>750,000</u>		<u>66,300</u>						0-99		
147636	<u>>750,000</u>	3,050	<u>281,000</u>	<400	<400	<u>503,000</u>					0-96	97-164	
141806	<u>>750,000</u>	7,000	<u>96,400</u>	<u>852,000</u>	<u>152,200</u>	<u>181,000</u>	<u>194,000</u>	<400	<400		0-96	97-164	165-272
134102	<u>535,000</u>	3,570	<u>>750,000</u>	40,200	<u>326,000</u>	<400	<u>88,400</u>		402	<u>7,440</u>	0-96	97-164	165-298

Table 11. Viral load histories of the ten children in this study. Plasma samples were taken at baseline and then every 4 weeks until week 24, then every 8 weeks until week 48, and then every 12 weeks until the end of the CHER trial. “*” Indicated those children who died by the end of the trial from non-AIDS related causes. Underlined viral loads indicated those samples that were used for sequence analysis in this study. See Table 12 for full details of the therapeutic history of each child.

3.1.1.2 Validation of cDNA synthesis and PR-RT amplification protocols

3.1.1.2.1 Sensitivity of protocols

The cDNA concentration that yielded 30% positive PCR reactions increased with increasing RNA inputs. The number of PR-RT sequences retrieved from HIV-1 single genomes for each viral load tested were 18 sequences for an input of 350 RNA molecules, 16 sequences for an input of 1,000 RNA molecules and 16 sequences for an input of 10,000 RNA molecules. 15-20 sequences derived from single genomes have an $\geq 80\%$ probability of detecting $\geq 10\%$ of the viral population [138]. The overall average pairwise nucleotide difference among the single genome derived PR-RT sequences was 0.0004 nucleotide substitutions per site with a standard deviation of 0.0001.

3.1.1.2.2 Specificity of protocols

All bulk amplicons derived from each patient sampling time point were determined to be HIV-1 PR-RT, from subtype C (Methodology in Section 2.2.7). A total of 1,026 PR-RT sequences derived from HIV-1 single genomes were analysed by maximum likelihood (ML) phylogenetic inference. This was because the sequences clustered together based on the patient that they were derived from (Figure 3.1). The same result was seen when ML trees included these sequences as well as those from TS5 and PR-RT the original p8MJ4-HSC and p8.9NSX+ plasmids (data not shown). TS5 sequences and plasmids PR-RT sequence also formed individual clusters

together in the ML tree. This data verified that there was no contamination between patient samples or from plasmids used in this project.

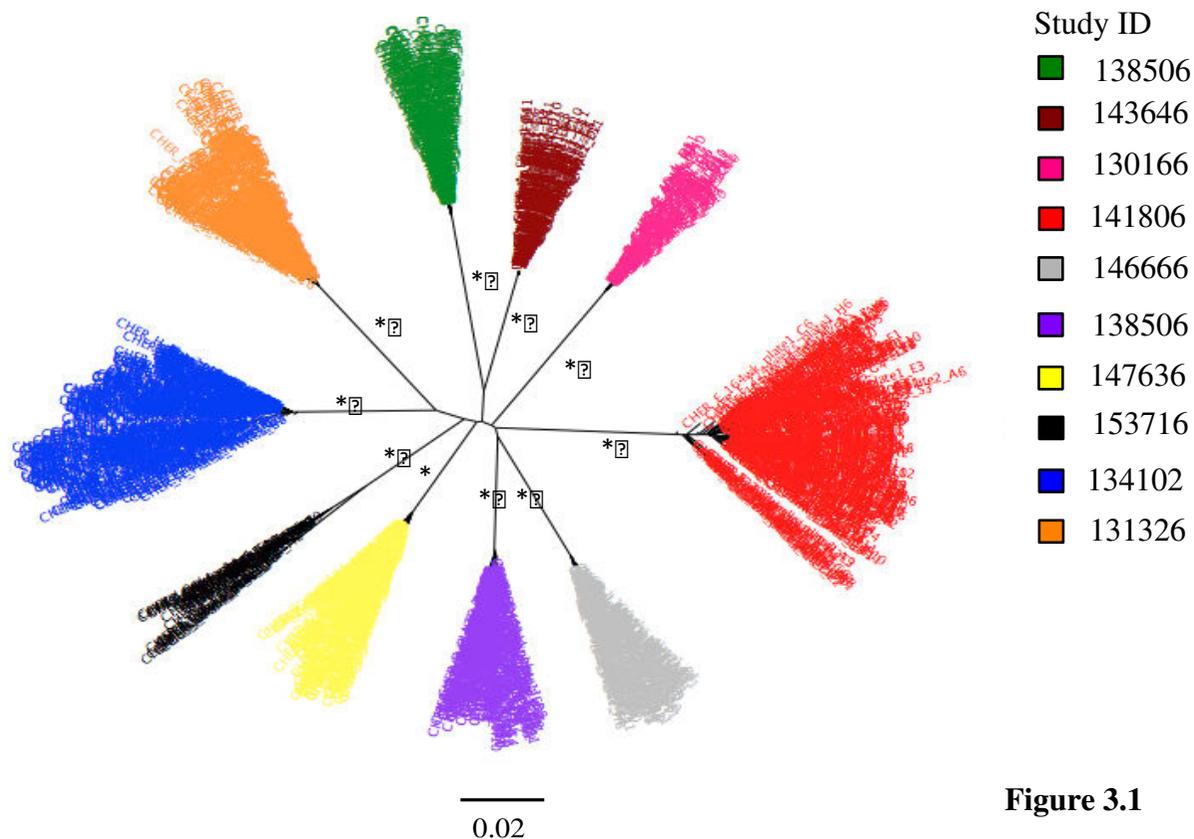


Figure 3.1 ML tree of 1,026 PR-RT single genomes sampled

from ten children from the CHER trial. Taxa in the tree colour-coded by Study ID. The scale is for the number of nucleotide substitution per site. * represents bootstrap values of 100.

3.1.1.3 PMTCT history of ten children in this study

8/10 children in this study cohort received perinatal NVP for PMTCT and all mothers received the same regimen at the onset of labour (Table 12).

Study ID	PMTCT NVP		Antiretroviral exposure
	Mother	Child	
143646	Yes	Yes	AZT, 3TC, LPV/r
141586	Yes	Yes	AZT, 3TC, LPV/r
130166	Yes	Yes	AZT, 3TC, LPV/r
131326	Yes	Yes	AZT, 3TC, LPV/r
153716*	Yes	Yes	AZT, 3TC, LPV/r
138506	Yes	Yes	AZT, 3TC, LPV/r
146666*	Yes	Yes	AZT, 3TC, LPV/r, RTV
147636	Yes	Unknown	AZT, 3TC, LPV/r, RTV
141806	Yes	Yes	AZT, 3TC, LPV/r
134102	Yes	No	AZT, 3TC, LPV/r, RTV

Table 12. CHER trial unique study identifications (Study IDs) allocated to the ten children in this study. The table also indicates their exposure to PMTCT/prophylactic NVP and antiretrovirals received during immediate therapy. * Indicates children who died of causes unrelated to AIDS before the end of the CHER trial.

3.1.1.4 Detection of baseline drug resistance

All PR-RT sequences analysed in this study were derived from HIV-1 subtype C. Bulk sequence analyses detected baseline NVP-selected resistance in 5/10 children: K103N (1/10), V106M (1/10), Y181C (2/10) and Y188C (1/10). However SGS detected NVP-selected resistance mutations in 7/10 (70%) of these children, and demonstrated a broader selection of mutations at higher frequency. The mutations detected by SGS were K103N (3/10), V106M (3/10) Y181C (3/10), Y188C (1/10),

A98G (1/10), K101E (2/10), V106A (2/10) and G190A (2/10). The results of baseline bulk sequencing and SGS for each child are shown in Table 13. All children had wildtype sequences present at a frequency >40%, however bulk sequencing was not able to detect these wildtype sequences in 3/10 children (Study IDs 130166, 153716 and 147636). A mixture of tyrosine (Y) and cysteine (C) were detected at position 181 in RT in the baseline viral population of Study ID 134102.

PI and NRTI DRMs were detected by SGS only and at baseline in two children, Study IDs 131326 and 153716. Study ID 153716 died before samples could be taken after 40 weeks of ART, but the major PI DRM, I50V was detected at a frequency of 3% (n=30 sequences) at baseline, which was not detected at week 40 of ART. T215I, a reversion mutation of the thymidine analogue mutation (TAM) T215F/Y [223, 224], was detected at a frequency of 5% (n=21 sequences) in the baseline viral population of Study ID 131326. This mutation was not detected at subsequent sampling time points, but virological failure occurred at week 40 with M46I in the minority quasispecies (3%, n = 38 sequences), This child achieved a viral load <400 c/ml by week 96 of early ART that was maintained until the end of the CHER trial.

Genetic linkage of RT DRM was detected in the viral population of two children. Y181C was linked with V106M (2%, n=49 sequences) and the TAM, K219N (2%) in the baseline viral population of Study ID 134102. Study ID 153716, also had a baseline minority variant with Y181C and K219N as genetically linked (3%, n=30 sequences) as well as a Y181C + L74V variant at a frequency of 7%..

Study ID	143646	141586	130166	131326	153716	138506	146666	147636	141806	134102
No of variants analysed (n)	49	30	33	21	30	57	26	34	30	50
Wildtype (%)	100	93	45	95	37	42	96	47	100	44
NNRTI resistance (%)										
A98G							4			
K101E			3					12		
K103N			36		7					4
V106A		7	3							
V106M						37		3		4
Y181C					40	21				44
Y188C			3					38		
G190A			9		3					
Y181C+V106M										2
NNRTI+NRTI resistance										
Y181C + L74V					7					
NNRTI resistance + TAM										
Y181C+K219N					3					2
TAM revertants										
T215I				5						
PI resistance										
I50V					3					

Table 13. Frequencies of baseline antiretroviral drug resistance variants by SGS. Frequencies in bold represent variants detected by both SGS and bulk sequence analysis.

3.1.1.5 Detection of low frequency NNRTI mutations during PI-based therapy

Despite treatment regimens lacking NNRTIs, NVP-selected mutations detected at baseline were also detected after 40 weeks of ART in two children by SGS (Table 14). In Study ID 147636, K101E was detected at week 40 of ART at a frequency of 7% but was not detected by bulk sequencing at either time point. In Study ID 134102, Y181C was detected at a frequency of 5% at week 40 of ART with SGS but was not detected by bulk sequence analysis at this time point. V108I was first detected at week 96 of ART at a frequency of 3%, then it was not detected at week 224 during re-started ART, and again detected at a frequency of 5% at week 298 during re-started ART. Emergence of V108I after ART initiation was not statistically supported ($p=0.189$) for test of proportions of genomes carrying V108I between baseline and week 298).

3.1.1.6 Detection of genetically linked dual-class resistance to PI and NRTI

Dual-class drug resistance was detected in the viral populations of 2 children (Study IDs 141806 and 134102) during ART by bulk sequence analysis. The same mutations were detected as linked on the same genome by SGS. M184V in RT and V82A in PR were detected on the same genome at weeks 40, 72 and 96 at frequencies of 68% ($n = 28$ sequences), 91% ($n = 44$ sequences), 79% ($n = 47$ sequences) respectively for Study ID 141806 (Table 15). This combination of genetically linked DRMs was also detected in the viral population of Study ID 134102 at weeks 96, 224 and 298 at frequencies of 29% ($n = 32$ sequences), 100% ($n = 34$ sequences), 100% ($n = 39$ sequences) respectively (Table 16). Two other dual-class drug resistant variants were

detected by SGS in the viral population of this child at week 96: M184V linked to M46I (31%) and M46L (3%) in PR (Table 16). Despite the high frequency of M46I, it was not detected by bulk sequence analysis. M46L could have been present at baseline given that only 1/32 sequences had this mutation at week 96 ($p = 0.457$, Fisher's Exact test of the proportion of sequences with M46L at baseline versus week 96).

Therapy status		Early ART		No ART	Re-started ART		
Week		0	40	96	164	224	298
Number of variants analysed for Study ID H (n)		34	29		39		
Number of variants analysed for Study ID J (n)		31	38	32		34	39
Study ID 147636	K101E (%)	12	7				
Study ID 134102	Y181C (%)	48	7				
	Y188C (%)		3				
	V108I (%)			3			5

Table 14. Frequencies of Nevirapine-selected NNRTI resistance mutations detected by SGS in two children, Study IDs 147636 and 134103.

The sampling time points for SGS are indicated. Frequencies in bold represent variants detected by both SGS and bulk sequence analysis.

3.1.1.7 Detection of genetically-linked triple class resistance to PI, NRTI and NNRTI

Genetically linked triple-class PI, NRTI and NNRTI DRMs were detected by SGS in Study ID 134102 during ART. Such variants were first detected at week 96 of early ART with additional RTV as a minority species (3%, n=32 sequences) The variant contained M184V, and V108I in RT and M46I in PR (Table 16). Variants with triple class DRMs were also detected at week 298 during re-started ART as a minority species (5%, n=39 sequences) and contained M184V, V108I and the PI DRMs: V82A, I54V, M46I, Q58E and L10F.

3.1.1.8 Multi-class drug resistance during ART is associated with high viral loads

M184V was the first DRM to be detected in the viral populations of Study IDs 141806 and 134102 who had virological failure with DRMs known to confer resistance to components of ART (Table 15 and Table 16). Viral load rebounds >50,000 copies/ml during ART in these children frequently coincided with dual-class drug resistance in the majority species (Table 11).

In Study ID 141806, 92% (n=37 sequences) of the viral population contained M184V only at week 12 of ART while the viral load fell from >750,000 copies/ml at baseline to 7,000 copies/ml at week 24 of ART. At week 40 of ART, there was a viral load rebound to 96,400 copies/ml. M184V single mutants were in the minority species

Therapy status	Early ART					No ART	Re-started ART
Week	0	12	40	72	96	164	224
Number of variants analysed (n)	33	37	28	44	47	38	22
Wildtype (%)	100	8				100	100
M184V (%)		92	32	9	21		
V82A + M184V (%)			68	91	79		

Table 15. Dual-class resistance detected by SGS during 1st line PI-based therapy for Study ID 141806. Frequencies in bold indicated variants detected by both SGS and bulk sequence. “+” Indicated linkage on the same genome. Early ART was stopped at week 96 and re-started at week 165.

(33%, n=28 sequences) of the viral population and the majority species (67%) contained M184V and V82A.

In Study ID 134102, the first viral load rebound was seen between week 24 and week 40 of early ART when the viral load rebounded from 3,570 copies/ml at week 24 of early ART to >750,000 copies/ml at week 40 of ART without the detection of multiple linked multi-class drug resistance; the quasispecies contained wildtype virus (47%), M184V (45%) and NNRTI DRMs Y181C and Y188C at a combined frequency of 8% (n=38 sequences). The viral load decreased to 40,200 copies/ml at week 72 of ART before rebounding to 326,000 copies/ml at week 96 of ART when receiving additional RTV, with dual-class drug resistance (M184V genetically linked to V82A or M46I) detected in the majority species (68%, n=32 sequences). The viral load was erratic thereafter, and at the last sample (week 298 during re-started ART) was measured at 7,440 copies/ml with dual-class drug resistance in 95% of the viral population (n=39 sequences), where these variants contained M184V genetically linked to 3 major and 2 minor PI resistance mutations (Table 16).

Therapy status	Early ART			Re-started ART	
Week	0	40	96	224	298
Number of variants analysed (n)	50	38	32	34	39
Frequencies (%)					
Wildtype	44	47			
M184V		45	28		
M184V + V82A			34	71	
M184V + M46I			31		
M184V + M46L			3		
M184V + M46I + V108I			3		
M184V + V82A + L10F				29	
M184V + V82A + I54V + L10F					3
M184V + V82A + I54V + M46I + Q58E + L10F					92
M184V + V82A + I54V + M46I + Q58E + L10F + V108I					5

Table 16. Dual- and triple-class resistance detected by SGS during 1st line PI-based therapy in Study ID 134102. Frequencies in bold indicate variants detected by both SGS and bulk sequence analysis. “+” Indicated linkage on the same genome. Early ART was stopped at week 96 and re-started at week 165.

3.1.2 Discussion

This study is the first to use SGS to characterise antiretroviral resistance in children on early combination ART. The extent to which differences in sequencing approaches impact the interpretation of drug resistance has been highlighted in previous studies [113-115]. In the present study, SGS was advantageous because it allowed a longitudinal assessment of the evolution of population diversity in each child and allowed us to determine genetic linkages of DRMs between viral PR and RT. However SGS remains unsuitable for routine clinical use because of the complexity of the methodology and interpretation of the data. I used Superscript III for cDNA synthesis (error rate = 1/30,000 nucleotides [225]) and Phusion High-Fidelity DNA Polymerase for PCR because of their low error rates (2.2% of PCR products would have 1 nucleotide mis-incorporation for a ~1.4kb product and 35 PCR cycles <http://www.thermoscientificbio.com/webtools/fidelity/>). Thus the fidelity of the PCR products was well maintained. Nonetheless, it is possible that some single instances of drug resistance mutations could be PCR artefacts.

SGS identified an additional 20% of children with NNRTI resistance as compared to bulk sequence analysis. SGS was also able to detect NNRTI resistance during PI-based ART in 2 of the children, which was not evident by bulk sequencing. I did not rule out the possibility that the DRMs I observed could have been the result of natural viral nucleotide variation.

At some of the time points in each child, bulk sequence analysis did not reveal any drug resistance, however SGS was able to detect DRMs in 25-35% of the sequences obtained. The limit of detection of bulk sequencing approaches to detect DRMs in *pol*

have been reported as 25%-35% [115] [226] [227] [228]. In my study, my SGS approach had a sensitivity that was comparable to that of Palmer et al [115], where any mutation detected at a frequency $\geq 35\%$ by SGS was detected with bulk sequence analysis. Similarly I saw a variation in sensitivity at some nucleotide positions. There were some mutations or wildtype codons that were detected at a frequency between 25% and 35% which were not detected by SGS. These were the wildtype amino acids in the baseline viral populations of Study IDs 130166, 138506 and 147636 and 82A and 46I, at frequencies of 31.2% and 29.4% respectively in the viral population at week 96 of ART for Study ID 134102. These discrepancies could be due to variations in primer sensitivities for their templates during cDNA synthesis and/or PCR amplification.

M184V was the first mutation selected in the majority of the viral population of the two children failing ART with DRMs; the viral load rebounds coincided with the majority of the viral population being replaced with dual-class drug resistant variants and notably no AZT-selected mutations were detected during ART. Furthermore, despite failing AZT and LPV/r-containing therapy, 215I and 219N in RT and I50V in PR were not detected by SGS or bulk sequencing during ART for the two children (Study IDs 131326 and 153716) that harboured these mutations in their baseline viral populations. These observations are consistent with previous data on the 'protective' effect of PI on the development of NRTI resistance [196, 229, 230].

Recently Palumbo et al [89] found that children given sdNVP for PMTCT were less likely to have treatment failure on LPV/r based first line therapy compared to NVP-based ART. This finding emphasizes the need for alternate PMTCT or ART strategies

in resource-limited settings. The WHO recommends first-line PI-based ART for such children and NNRTI-based therapy as second-line therapy, usually NVP or EFV. But very few studies have described the persistence of NNRTI resistance selected by prophylactic single dose NVP during PI-based ART in vertically infected children.

Bulk sequence analysis did not reveal persistence of NNRTI resistance during PI-based ART, but SGS revealed persistence of such mutations in more than one child as minority species. Prophylactic NVP was not given to 134102, but this child's mother received it at the onset of labour, 7.9 weeks before the baseline sample was taken (Appendix A Table 21). I do not know exactly when maternal sdNVP was given relative to birth, but the current literature concludes that maternal NVP can remain at a dose representing 10 times the IC₅₀ in neonates for 2.4 weeks or at suboptimal levels for 4 weeks after sdNVP is given to the mother at the onset of labour (30 to 480 minutes prior to birth) [231]. Therefore, assuming that this child was not breast-fed, it is possible that maternal NVP was transferred via the placenta, which selected the variety of NNRTI resistant viruses detected at baseline in this child: K103N, V106M and Y181C. Alternatively, the presence of NNRTI resistance in this child without a history of sdNVP for PMTCT was most likely explained by vertical transmission of DRMs[119] or maternal NVP [231].

The baseline PI DRM I50V, TAM revertant T215I and the TAM K219N of the double mutant Y181C + K219N, that were detected in Study IDs 131326, 153716 and 134102 may have also been vertically transmitted. Vertical transmission of drug resistance has been previously determined by phylogenetic comparison of maternal

and neonatal sequences[232]. Unfortunately, maternal plasma samples were not available for my study.

Age adjusted, full-dose RTV can select M46I, I54V and V82A in children [233], as well as L10F, M46L, Q58E in adults[234]. I detected genetic linkage of all these mutations in the viral population of 134102 after RTV super-boosted LPV/r, conferring greater PI resistance and reducing treatment options as compared to Study ID 141806 [102]. Currently there are no published data on the selection of resistance after RTV super-boosting of LPV/r, therefore my findings, although novel and important, require verification in a larger sample cohort.

The current virological understanding of drug resistant reservoirs suggests that multi-class drug resistant viruses can compromise ART in the future, although this requires formal testing in children where drug class options are limited. I consider the possibility that V108I was selected by NVP for PMTCT because it is not usually found in patients infected with subtype C who were not exposed to this drug [102]. My evidence also points to the triple class drug resistant variants detected in Study ID 134102 to be a result of genetic 'hitch hiking' by V108I. V108I is known to confer modest reductions in NVP/EFV susceptibility *in vitro* [102] but its clinical significance is not known [102].

Although multi-class drug resistance may be more common than previously thought, it should be emphasised that the CHER trial demonstrated excellent outcomes overall with 84.6% (280/331) of the children on ART achieving viral loads <400 c/ml at the end of the trial. These include 2/10 children from my study cohort (Study IDs 130166 and 141806) who achieved and maintained viral loads <400 copies/ml at week 164

and week 265 respectively, until the end of the CHER trial. Further large-scale studies are needed to address long-term implications in these NVP exposed, vertically infected children with virological failure on LPV/r, and in particular the efficacy of the second-generation NNRTI etravirine and the PI darunavir warrant investigation. Finally these data reinforce the urgent requirement for appropriate paediatric formulations of a wider array of antiretroviral medications in the future.

3.1.3 Conclusions

Although the CHER trial demonstrated excellent outcomes overall with 84.6% (280/331) of the children on ART achieving viral loads <400 c/ml at the end of the follow up period, my results show the potential for early ART in vertically infected children to give rise to multi-class resistant viruses after NVP-based PMTCT and persistence of NNRTI resistance during ART. These findings continue to suggest a clinical role for highly sensitive assays like SGS to detect and quantify drug resistant viruses at low frequencies and to determine the inter- and intra-host prevalence of multi-class drug resistance.

I also reinforce previous conclusions that RTV super-boosting of LPV in infants and children poses a significant risk of development of multiple protease inhibitor resistance conferring mutations, not just within the viral population but within single virions as majority species. This further highlights the urgent need for appropriate paediatric formulations of a wider array of antiretroviral medications in the future.

The current virological understanding of drug resistant reservoirs suggests that multi-resistant HIV will compromise ART in the future, although this requires formal

testing in children. These data highlight future therapeutic challenges as a large cohort of vertically infected children on ART move into adolescence.

In the following chapters I continue to address the evolution of drug resistance in these children using phylogenetic and phenotypic techniques.

3.2 Phylogenetic analysis of PR-RT to characterise the development of multiple linked multi-class drug resistance

This chapter addresses the second objective of this thesis: to characterise the evolution of the viral populations using phylogenetic tools for the children from my study cohort. This was done particularly given of the development of multiple linked multi-class drug resistance. I also determined the evolutionary profile of NVP selected resistance that was detected at baseline and during early ART. Finally, I wished to determine if drug resistant and drug susceptible viruses, that were detected after and before early ART started persisting from previous time points or emerged from independent evolutionary events. Data for Study ID 153716 was not included in these analyses because I only obtained 1 sequence from one of the two sampling time points for this child.

Genetically linked PI and RTI resistance mutations developed in the majority species of the viral populations of two children, Study IDs 141806 and 134102, and were detected as genetically linked by SGS. Previously, I saw that these variants were detected in the majority species during the first-round of ART and during re-started ART for Study ID 134102. For Study ID 141806, multi-class drug resistant variants were detected in the majority species during the first-round of ART, but were not detected during re-initiated ART. Instead, sequences without DRMs re-emerged during re-initiated ART and persisted as the majority species.

Sequences with G-to-A mutations due to APOBEC3G/F cytidine deamination would have confounded the diversity calculations in the viral populations of the children from our study cohort. Therefore I submitted MSAs of PR-RT derived single genomes

obtained for each child at all sampling time points to the Los Alamos Hypermut

2.2 algorithm:

<http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html>. Results of this analysis can be viewed in Appendix B.

Correlations between the genetic variation of PR-RT and viral load trends, or between genetic variation and development of viruses with multiple drug resistance conferring mutations linked on the same genome were determined. I used the Molecular Evolutionary Genetic Analysis (MEGA) software version 5.2 [217] to calculate the mean pairwise genetic distances (MPDs) of PR-RT sequences from each sampling time point per child (Appendix B Table 23). In this case, I used the Tamura and Nei 1993 nucleotide substitution model, which was determined as the best fit model for this data by MEGA (lowest Bayesian Information Criterion, therefore highest posterior probability). Regression analysis was used to determine if the number of sequences obtained at each time point affected nucleotide variation. I determined whether differences in the mean number of nucleotide substitutions per site in PR-RT MPDs between consecutive time points were significant using an unpaired two-tailed Student's t-test. MPD was expressed as number of nucleotide substitutions per site of all PR-RT sequences. For 4/9 children (Study IDs 138506, 141586, 143646 and 130166), there were not enough data points to do this analysis.

Next, I divided the children into 3 groups based on their ART experience: Group 1 had early ART until week 40 or 96 and they did not re-start ART during the CHER trial (n = 4 children; Study IDs 141586, 138506 and 147636). Since the last sampling time point before death for Study ID 146666 was at week 99 of early ART, this child

was included in Group 1. Group 2 received early ART that was uninterrupted until the end of the CHER study (i.e. continuous ART; n = 3 children; Study IDs 143646, 131326 and 130166). Group 3 received early ART which was interrupted from weeks 97 to 164 and then the same ART regime was re-started from week 165 until the end of the CHER trial (i.e. interrupted ART; n = 2 children; Study IDs 141806 and 134102). I plotted both viral load (copies/ml) and the mean pairwise genetic distances (nucleotide substitutions per site in PR-RT) against sampling time (weeks since start of early ART) for these children.

The evolution of drug resistance was depicted from maximum likelihood inferences for each child. I re-constructed viral phylogenies for each child using the program RaxMLGUI [235]. The parameters used for these analyses were the general time reversible nucleotide substitution model, rapid bootstrap heuristics to do 1000 independent RaxML searches to choose a tree with optimal ML value and to assess branch support with rapid bootstrapping. An HIV-1 subtype C PR-RT single genome sequence derived from the test sample was used as an outgroup to root the trees to assume the direction of evolution.

I implemented recombination analyses to determine if recombination was a mechanism by which multiple linked multi-class drug resistant variants emerged from the viral populations of Study IDs 134102. I looked for evidence of recombination using SBP [236] and GARD [237] provided online at <http://www.datamonkey.org>.

Codon-based maximum likelihood inferences were used to identify positively selected residues in the viral populations of Study IDs 141806 and 134102. The three methods used were: (1) single ancestor counting (SLAC) analysis, (2) fixed effect likelihood

(FEL) analysis [218, 238] to estimate dN/dS ratios at every codon in the MSA for each child and (3) Fast Unconstrained Bayesian AppRoximation (FUBAR) [220]. Spidermonkey analysis was used to determine if positively selected sites co-evolved, i.e. if the evolution of amino acids at any pair of positively selected sites were dependent on each other during the course of protein evolution. MSAs were submitted to the SLAC, FEL, FUBAR and Spidermonkey algorithms at Datamonkey webserver [221] at <http://www.datamonkey.org>.

3.2.1 Results

3.2.1.1 Correlations between the number of single genome sequences obtained at each time point and nucleotide variation

There was no correlation between the number of single genomes obtained at each sampling time point and the genetic variation of PR-RT ($p > 0.05$). This suggested that the amount of genetic variation seen in the viral populations over time reflects the real diversity of the population and is not influenced by sample size. Figure 3.2 depicts the linear regression analyses for these children.

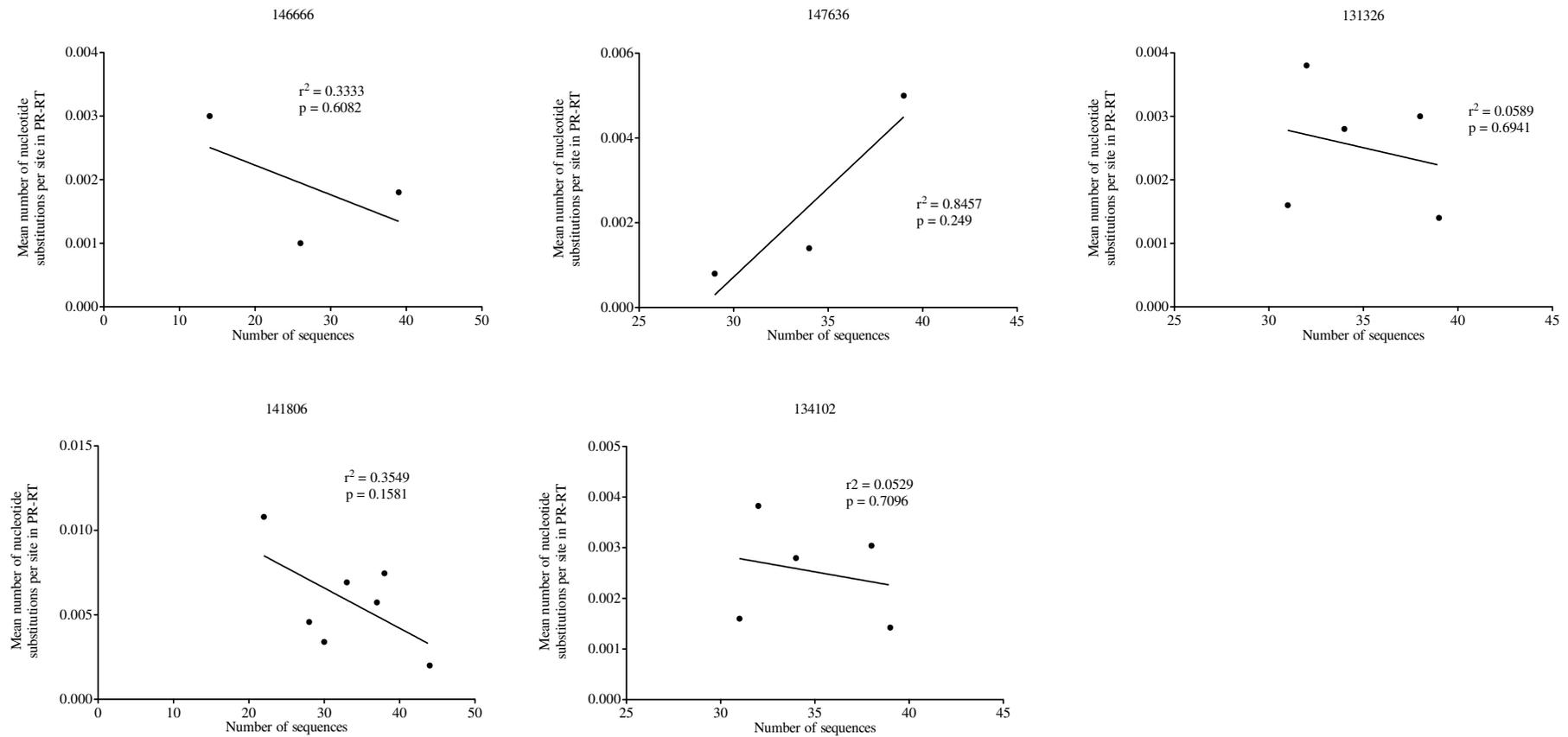


Figure 3.2 Regression analysis for 5/9 children. r^2 and p -value of the F-test determine if the slope of the best-fit line is true. When $p > 0.05$, the slope of the line was not true and there was no correlation between number of single genomes obtained and MPD.

3.2.1.2 Effect of ART on PR-RT genetic diversity

None of the children from Group 1 maintained a viral load that was below the limit of detection (<400 copies/ml) even though 2 children achieved this during therapy. All the members of Group 2 achieved undetectable viral loads during ART at time points later than week 40. They maintained undetectable viral loads until the last sampling time point during the CHER trial. The members of Group 3 achieved undetectable viral loads during the first round of ART and only Study ID 141806 achieved an undetectable viral load during the second round of ART, which was maintained until the end of the CHER trial.

In Group 1, 1/4 children (Study ID 138506) received ART until week 40 and 3/4 children (Study IDs 141586, 146666 and 147636) had ART until week 96. The diversity of the viral population at baseline and week 40 of ART remained similar for Study ID 141586 ($p > 0.05$; Student's t-test of the difference between MPDs at different sampling time points). For 2/4 children, (Study IDs 138506 and 147636) the diversity of viral populations at week 40 of ART was greater than at baseline. For the fourth child in this group (Study ID 146666), the diversity of the viral population showed no particular pattern. These longitudinal trends in viral load and diversity versus time are shown in Figure 3.3A.

In Group 2, children received continuous ART and the diversity of the viral populations at week 40 of ART was less than at baseline for 2/3 children (Study IDs 143646 and 130166), whereas this trend in population diversity was not seen for the

third child in this group (Study ID 131326). These longitudinal trends in viral load and diversity versus time are shown in Figure 3.3B.

In Group 3, early ART was interrupted from weeks 97-164. For Study ID 141806, the diversity of the viral population was less than that of the previous time point until week 72 of ART and then was greater at each subsequent time point until week 224.

The opposite trend in population diversity observed For Study ID 134102: the diversity of the viral population was greater than the last until week 96 of early ART and then was less diverse at each subsequent time point until week 298. These longitudinal trends in viral load and diversity versus time are shown in Figure 3.3C.

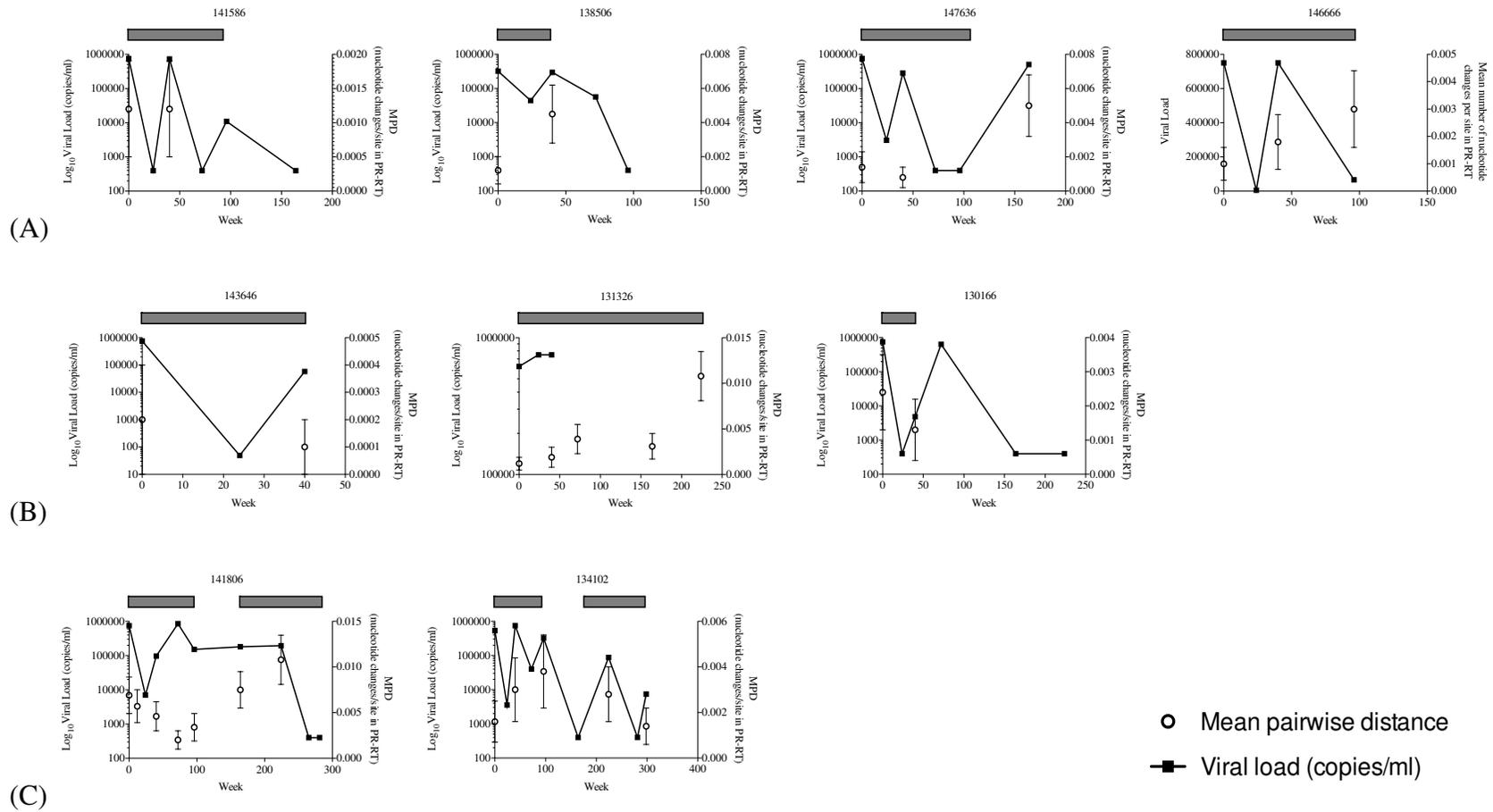


Figure 3.3 Longitudinal HIV-1 plasma viral load (copies/ml) and diversity (MPD/nucleotide changes per site in PR/RT) profiles for children from (A) Group 1, (B) Group 2 and (C) Group 3. Study IDs are above each graph. Grey bars represent time on ART.

Within each group, there was no difference between MPDs before early ART was started compared to when the first viral rebound was observed during early ART in all three groups ($p > 0.05$; Student's t-test). There was also no difference among the baseline MPDs of each group or among the MPDs of each group at week 40 of ART ($p > 0.05$; One-Way ANOVA test) (Figure 3.4).

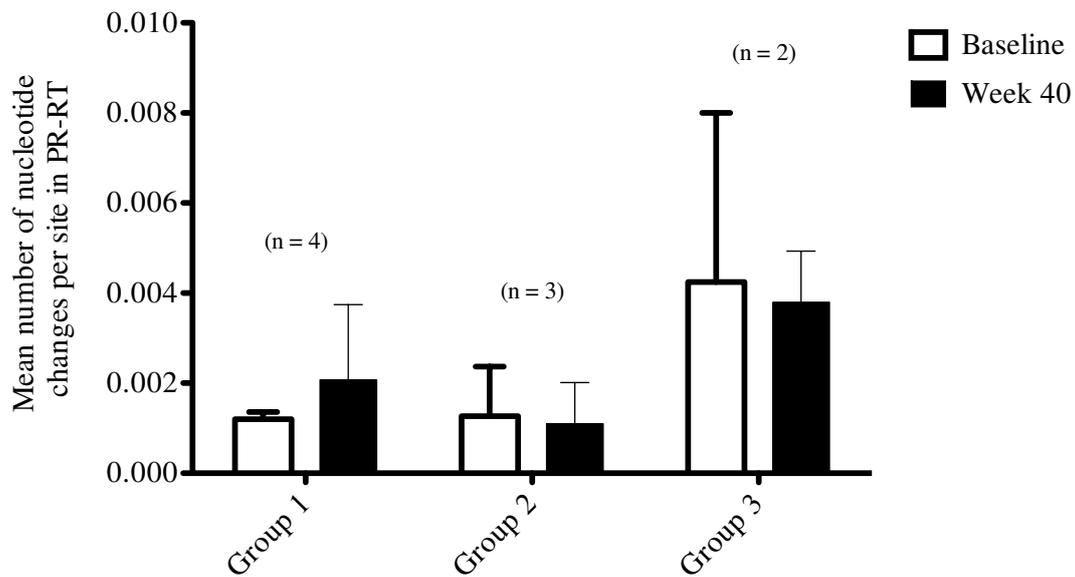


Figure 3.4 PR-RT diversity calculated as MPD before early ART (Baseline) and at the time of the first viral rebound during ART (Week 40) for 9 children. Group 1 contains children who received early ART until week 40 of 96. Group 2 contains children who received continuous ART and Group 3 contains children who received interrupted ART. The number of children (n) in each group is shown in parentheses above each bar couple. Error bars represent the standard deviation of mean MPDs.

3.2.1.3 Effect of ART on phylogenetic structures

The ML trees reconstructed with the sequences obtained from each child reflected the longitudinal trends in diversity I saw in each patient. My data showed 6 children from Groups 1 and 2 that had clusters of intermingled, identical sequences from baseline and week 40 of ART (Group 1: Study IDs 138506, 146666 and 141586 – (Figures 3.5, 3.6, 3.7 and 3.8; Group 2: Study IDs 143646, 131326 and 130166 – Figure 3.9, 3.10 and 3.11). This observation was over-represented for one patient who received continuous ART (Study ID 143646, Figure 3.9); the majority (42/49) of sequences obtained from baseline and week 40 of ART were identical.

I also observed clusters of identical sequences from different time points when a drug resistant population was selected by components of ART in one child (Study ID 141806; Figure 3.12) from Group 3. Such a cluster of sequences existed in the major M184V single-mutated population and the major M184V+V82A double-mutated population. Both of these populations shared a common ancestor and sequences were very closely related when they were not identical. The major M184V single-mutated population shared a common ancestor with the major double-mutated M184V+V82A population and V82A single mutated sequences were never detected in the viral populations of this child.

The other child in Group 3 (Study ID 134102; Figure 3.13) had identical sequences from the same time point and with the same drug resistance mutation profiles. M184V single mutated viruses were selected from multiple viral populations but only one population was the major contributor to drug resistant viraemia at week 40 of ART.

At week 96 of ART, there were 2 main viral populations that were the major

contributors to the viraemia. One was an M184V single-mutated population, from which a double-mutated M184V+M46I viral population emerged. The other population was a M184V+V82A double-mutated population, which shared a common ancestor with the major M184V single-mutated viruses from week 40 and 96 of ART. When ART was re-started in this child, it was the M184V+V82A double mutated viral population from week 96 of ART that was ancestral to the multiple-linked multiclass drug resistant populations detected at weeks 224 and 298. At week 224 during re-started ART, all viruses were double-mutated with M184V and V82A and with the accumulation of LPV-selected mutations in PR (L10F, M46I, I54V Q58E) by week 298 of ART, the viral population became more homogenous.

For Study ID 134102, despite a viral load <400 copies/ml at the end of treatment interruption (week 164) the drug resistant population at week 224 was not independently selected by re-started ART, but instead evolved from the major M184V+V82A sub-population at week 96 of ART. Despite the suppression of viral replication to 402 copies/ml at week 265 of re-started ART, the multiple linked multi-class drug resistant population at week 298 during re-started ART evolved from the population at week 224 instead of being independently selected. In contrast, after ART was interrupted and re-started for Study ID 141806, no drug resistance mutations were detected. When ART was re-started, the viral load rebounded from undetectable (<400 copies/ml at week 164) to detectable (88, 400 copies/ml at week 224) in the absence of known drug resistance mutations in PR or RT.

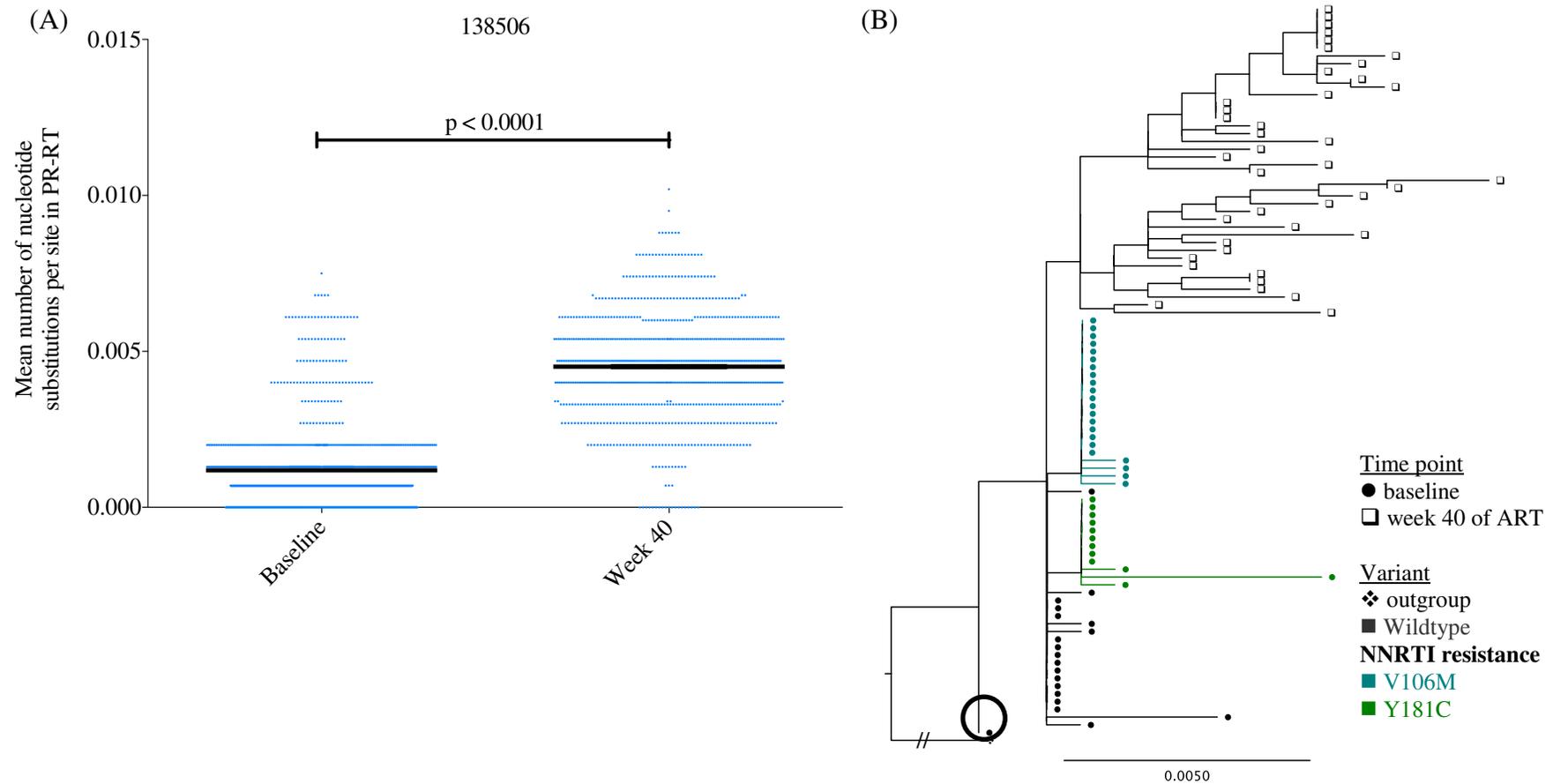


Figure 3.5 (A) MPD and (B) Phylogenetic structure of sequences from Study ID 138506. Both contain PR-RT sequences from baseline and Week 40 of ART. Encircled is the most ancestral sequence.

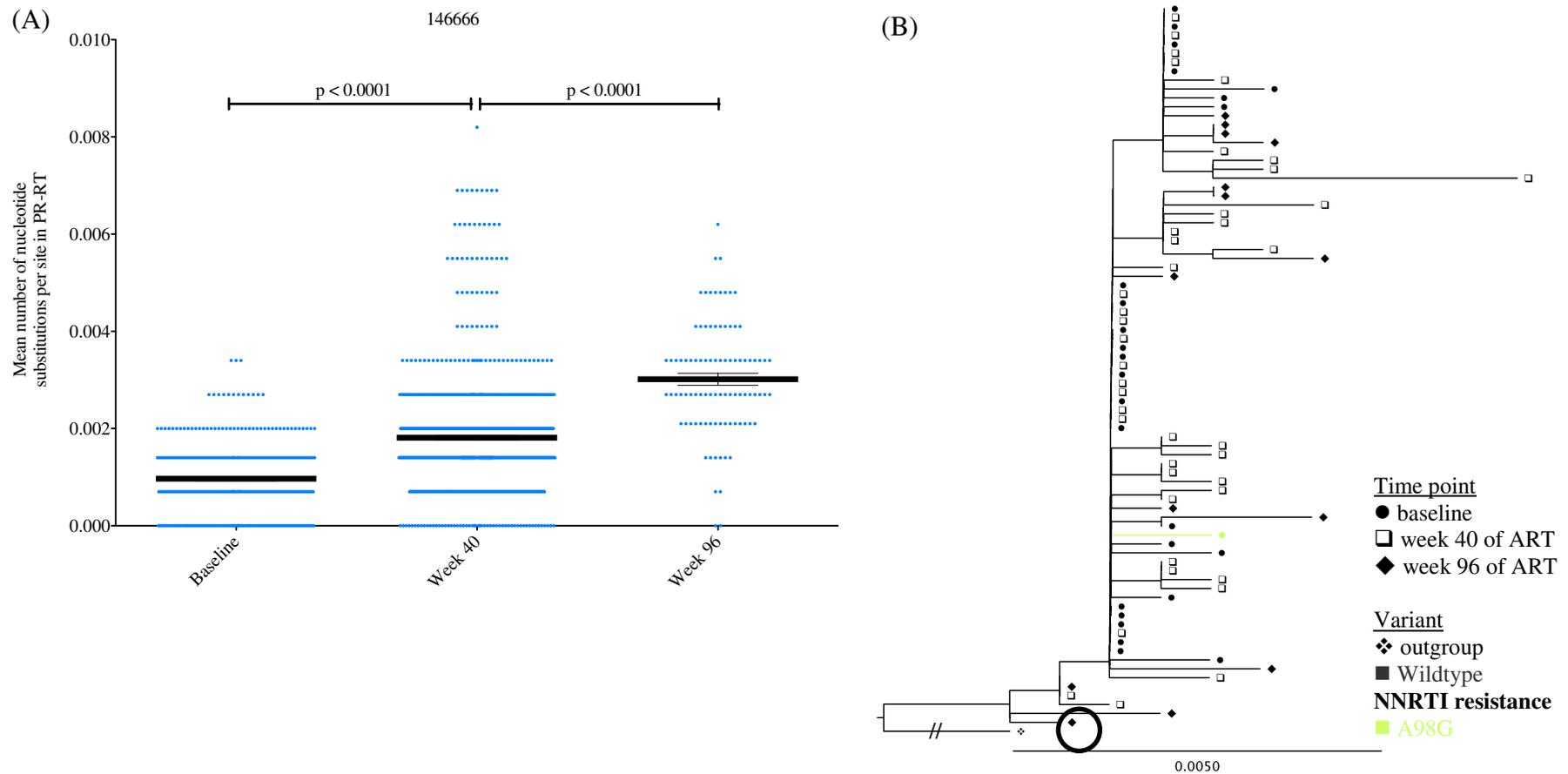


Figure 3.6 (A) MPD and (B) Phylogenetic structure of sequences from Study ID 146666. Both contain PR-RT sequences from baseline, Week 40 and 96 of ART. Encircled is the most ancestral sequence.

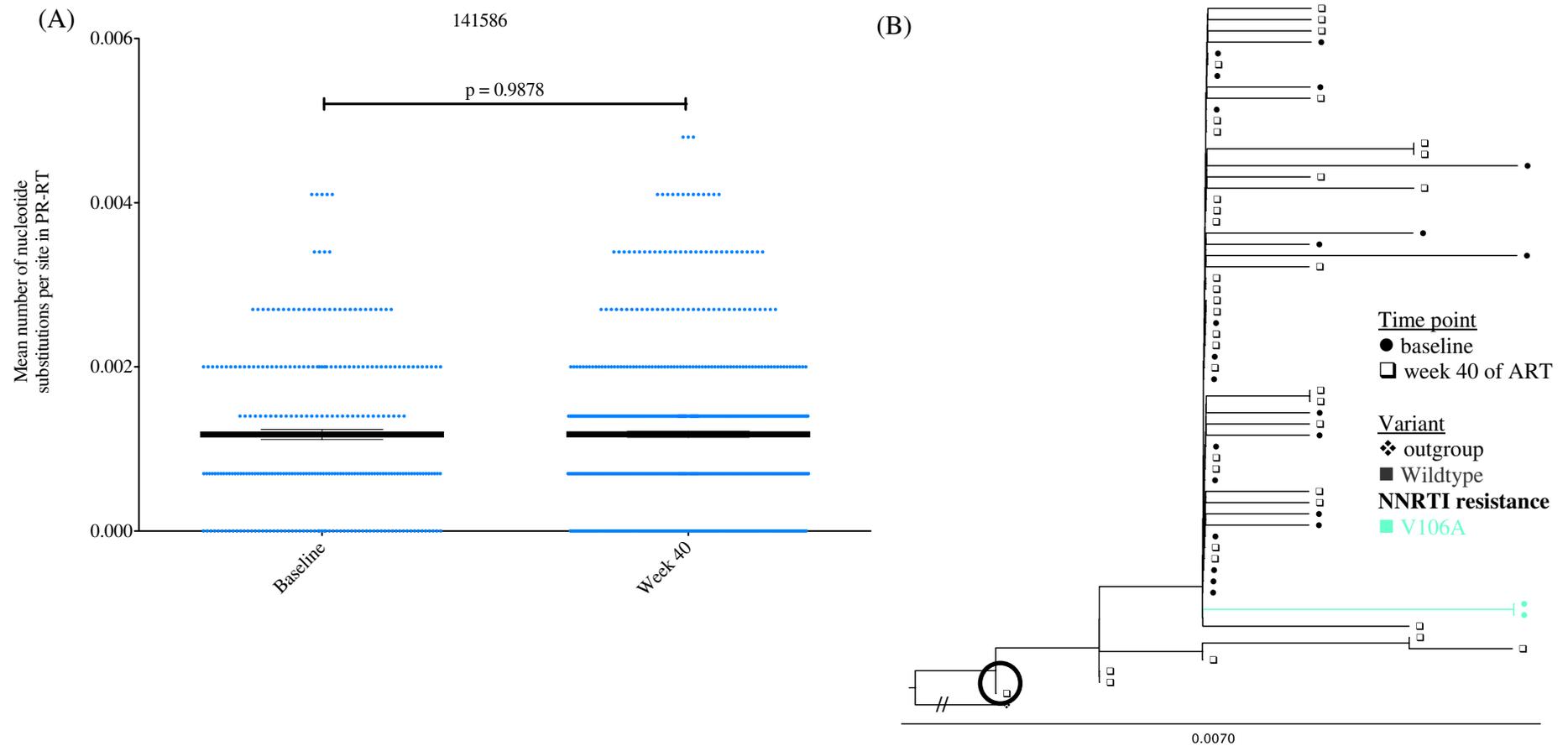


Figure 3.7 (A) MPD and (B) Phylogenetic structure of sequences from Study ID 141586. Both contain PR-RT sequences from baseline and Week 40 of ART. Encircled is the most ancestral sequence.

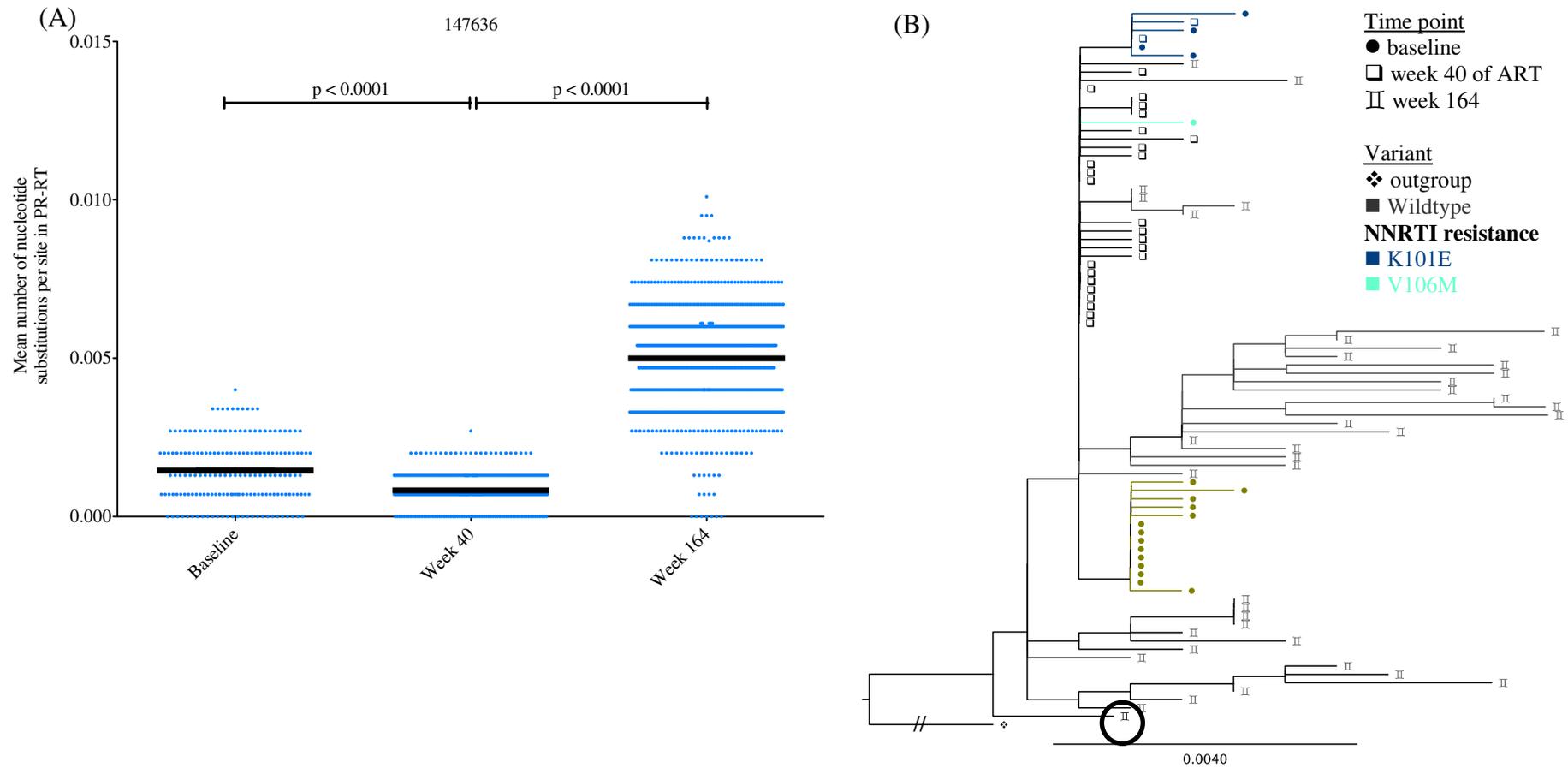


Figure 3.8 (A) MPD and (B) Phylogenetic structure of sequences from Study ID 147636. Both contain PR-RT sequences from baseline, Week 40 of ART and Week 164 without ART. Encircled is the most ancestral sequence.

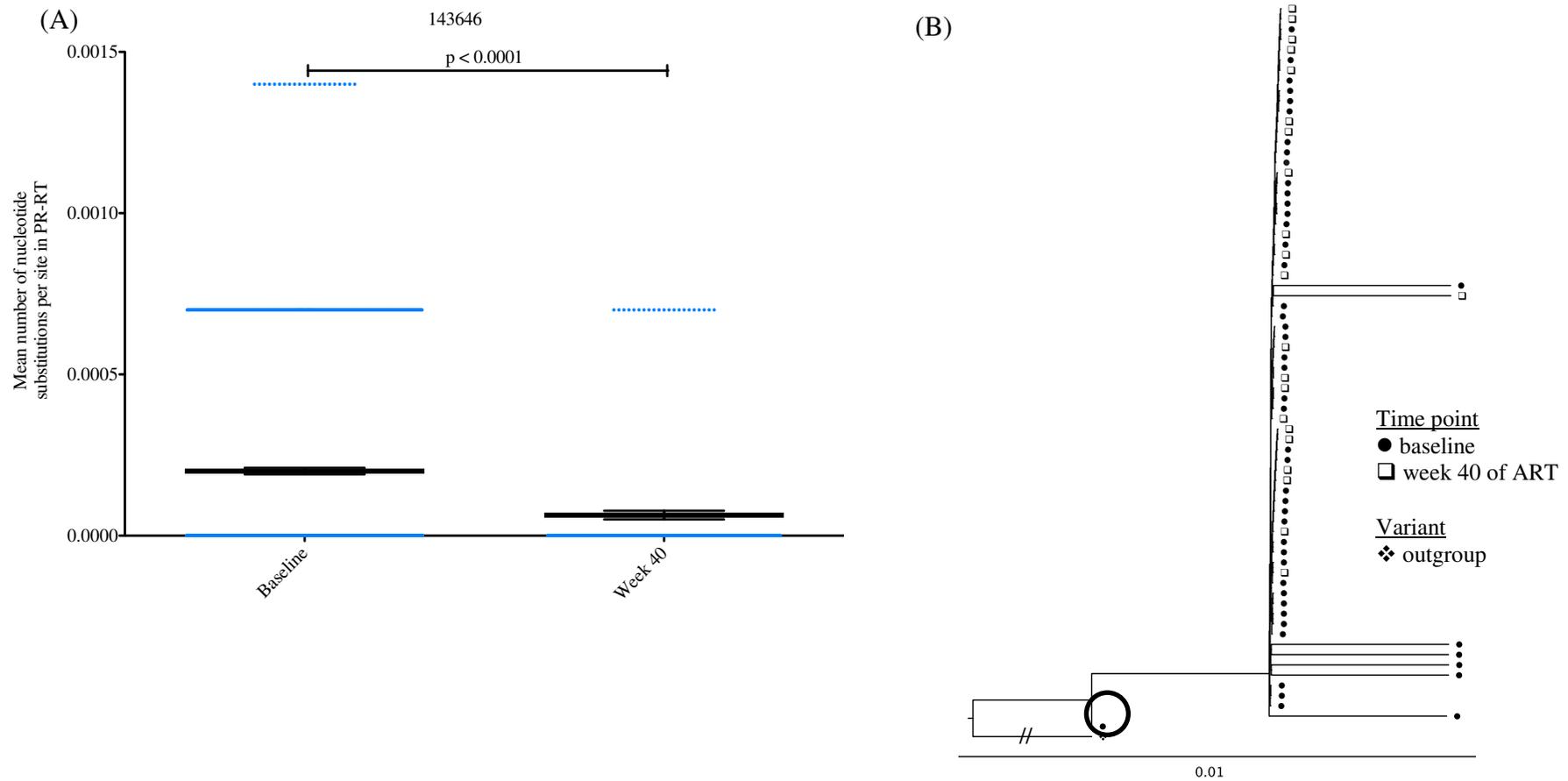


Figure 3.9 (A) MPD and (B) Phylogenetic structure of sequences from Study ID 143646. Both contain PR-RT sequences from baseline and Week 40 of ART. Encircled is the most ancestral sequence.

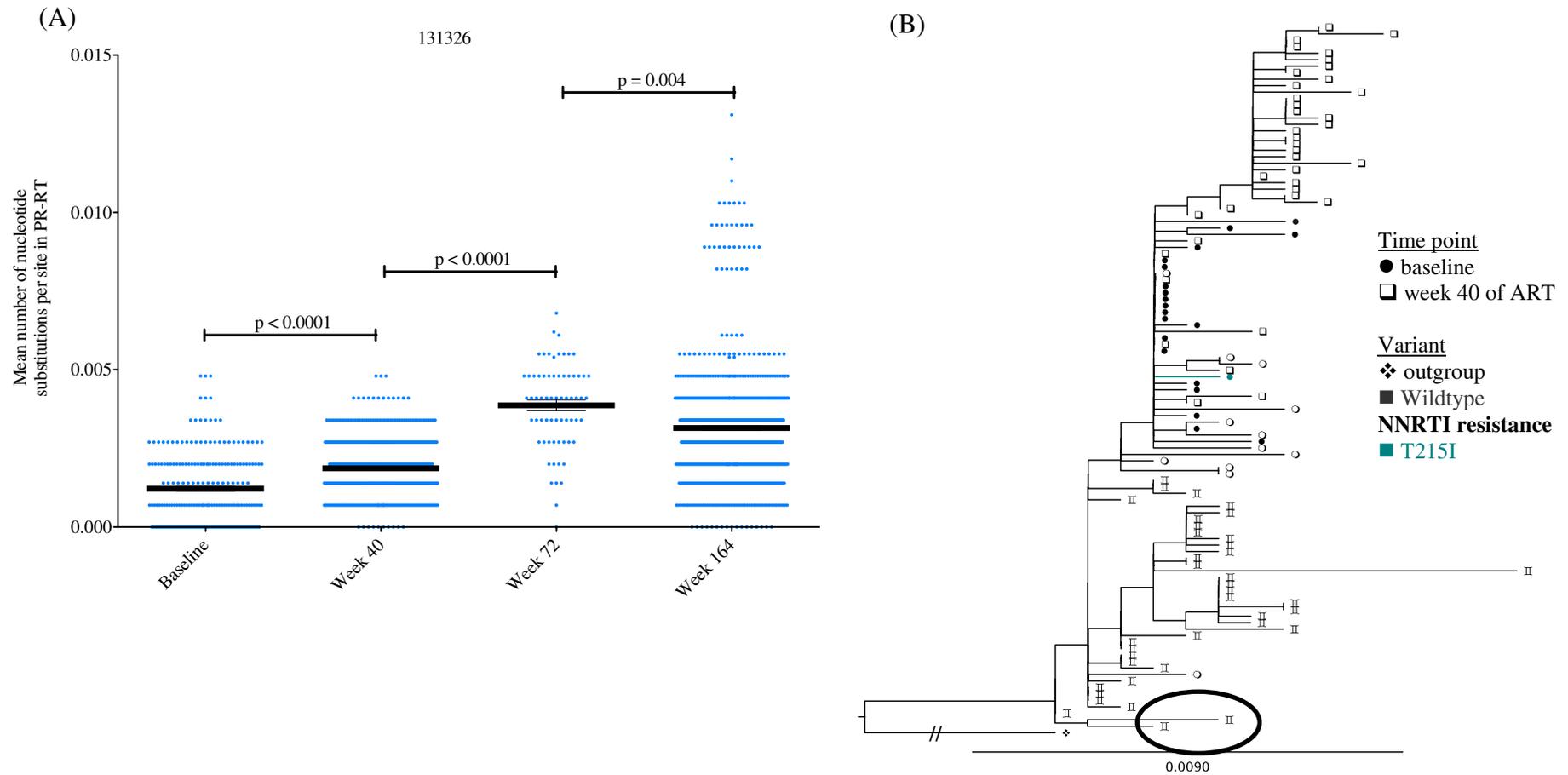


Figure 3.10 (A) MPD and (B) Phylogenetic structure of sequences from Study ID 131326. Both contain PR-RT sequences from baseline and Weeks 40, 72 and 164 of ART. Encircled is the most ancestral sequence.

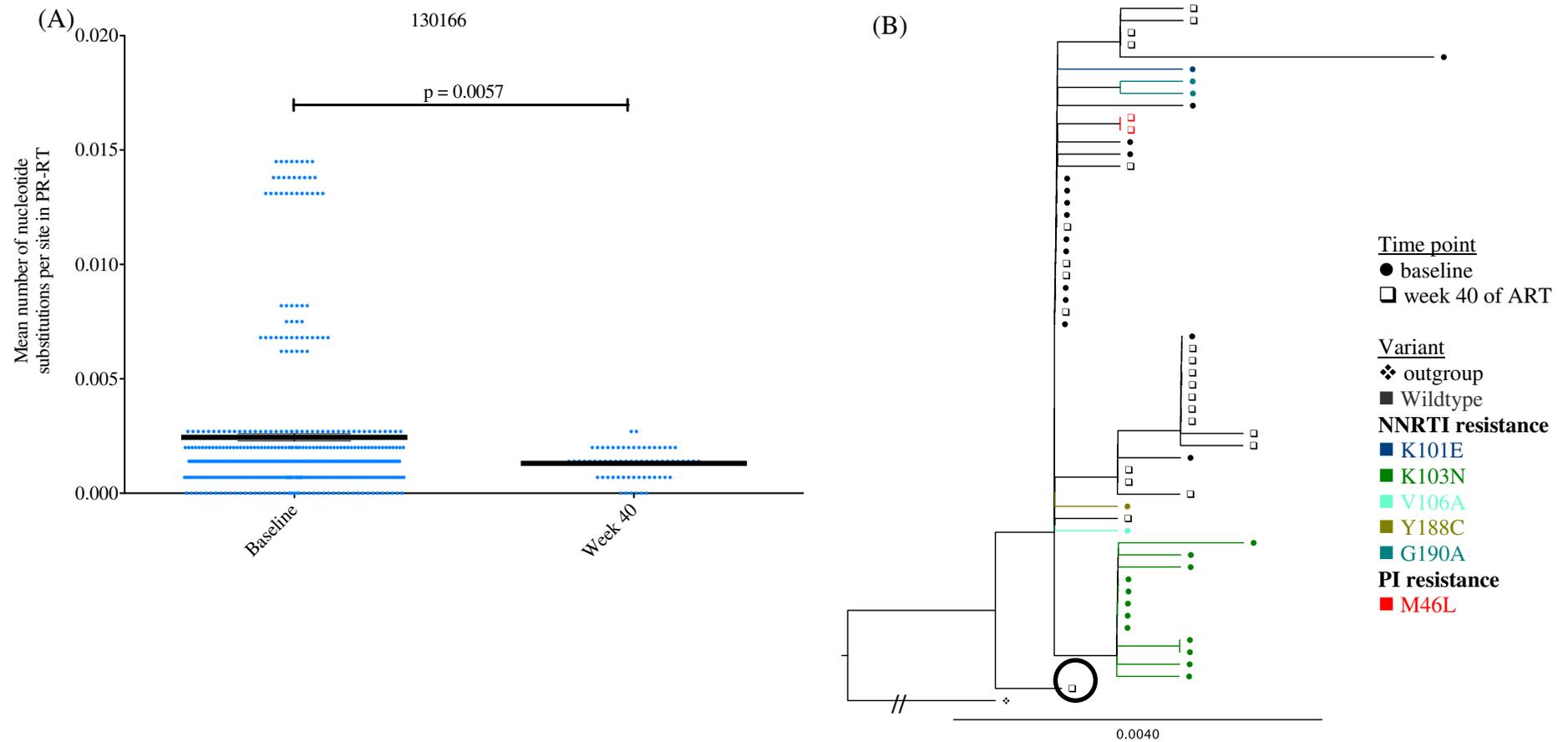


Figure 3.11 (A) MPD and (B) Phylogenetic structure of sequences from Study ID 130166. Both contain PR-RT sequences from baseline and Week 40 of ART. Encircled is the most ancestral sequence.

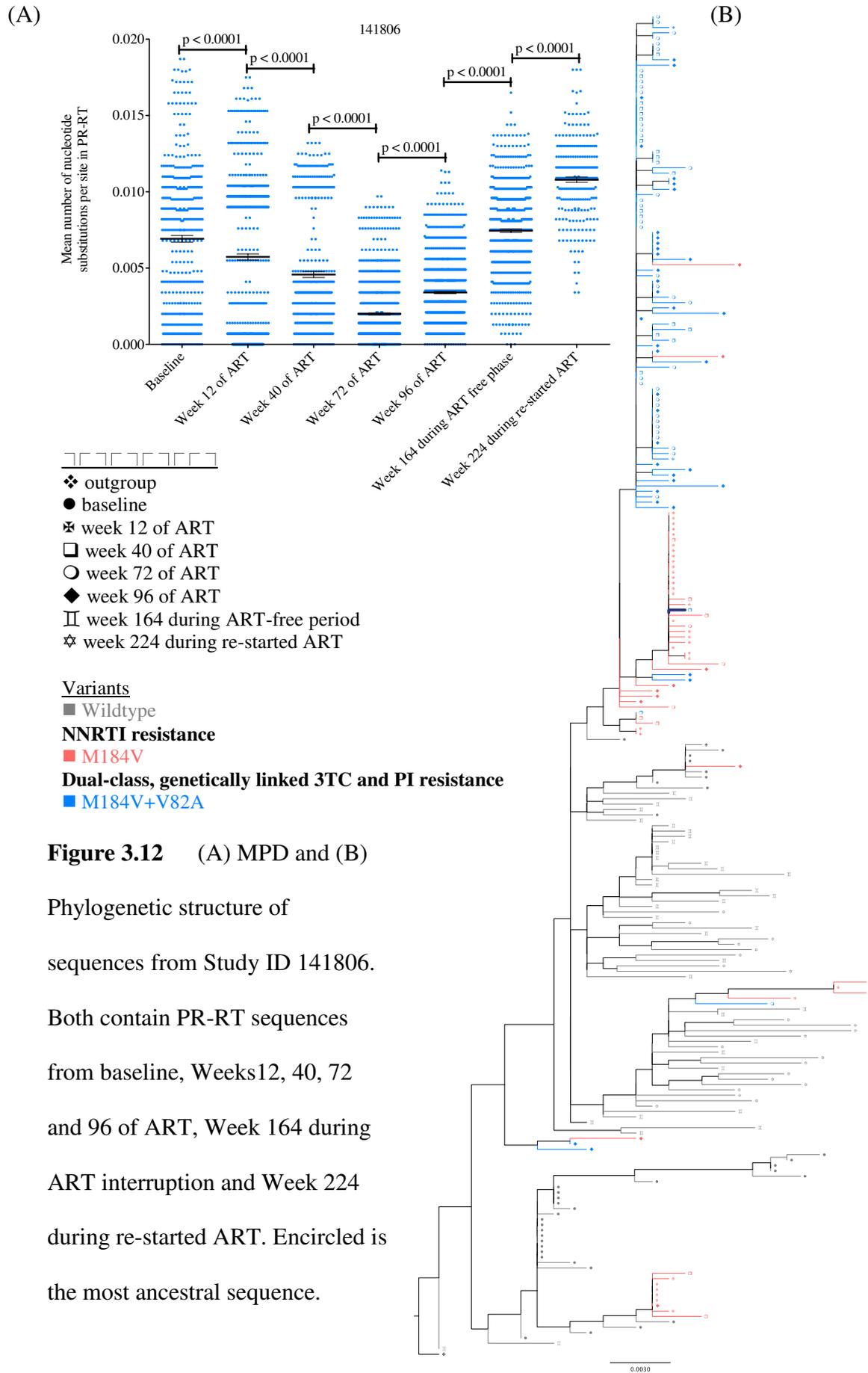


Figure 3.12 (A) MPD and (B) Phylogenetic structure of sequences from Study ID 141806. Both contain PR-RT sequences from baseline, Weeks 12, 40, 72 and 96 of ART, Week 164 during ART interruption and Week 224 during re-started ART. Encircled is the most ancestral sequence.

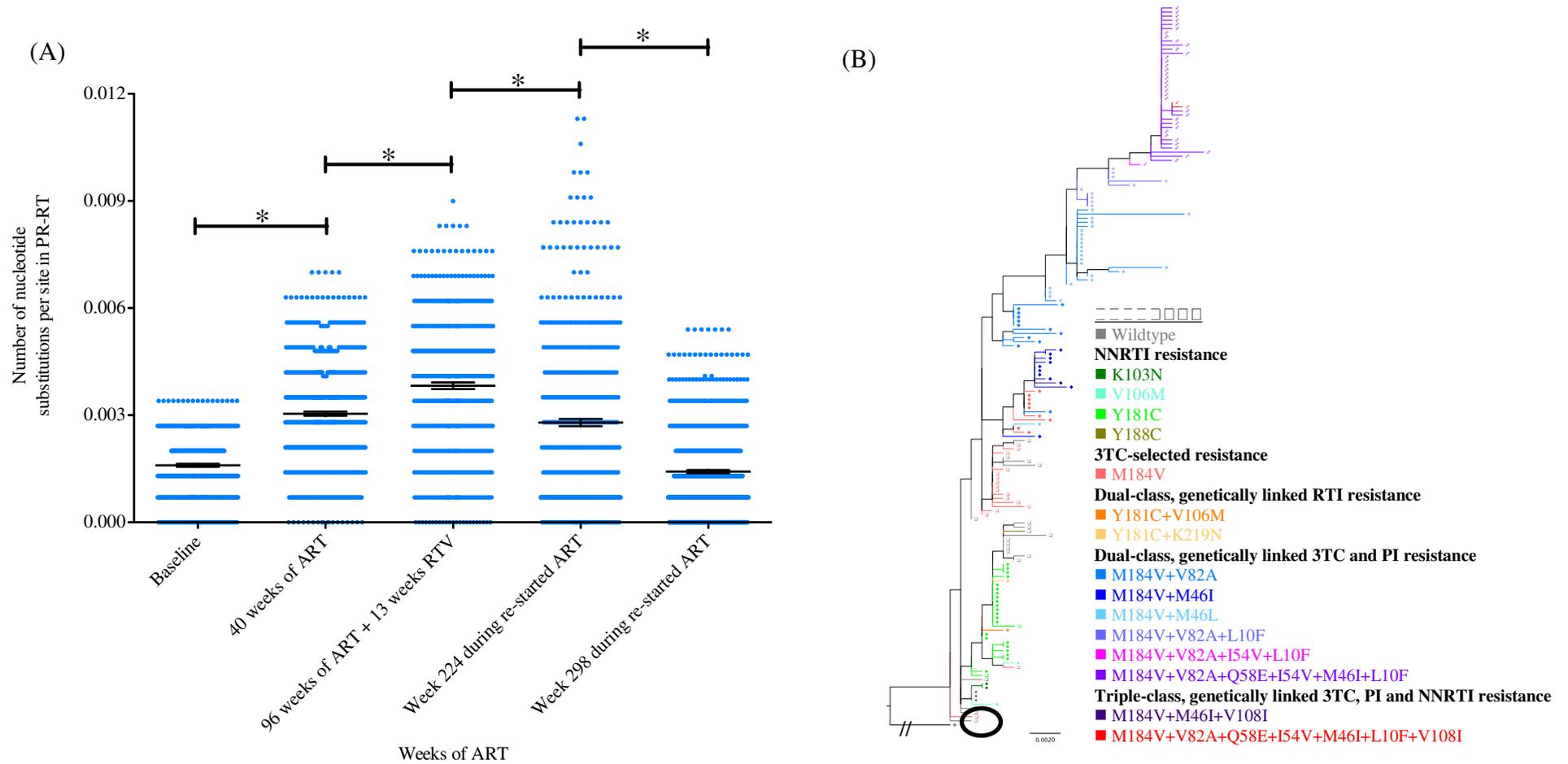


Figure 3.13 (A) MDP and (B) Phylogenetic structure of sequences from Study ID 134102. Both contain PR-RT sequences from baseline and Weeks 40 and 96 of ART, Weeks 224 and 298 during re-started ART. Encircled is the most ancestral sequence.

3.2.1.4 Ancestry of the viral population during early ART

Since the rooted ML trees allowed the assumption of the direction of evolution, I was able to observe the most ancestral sequence of all the single genome derived PR-RT sequences as the sequence closest to the base of the tree. For 9/9 children, the most ancestral sequence did not contain a known DRM in PR or RT. For 2/9 children (Study IDs 138506 and 143646; Figures 3.5 and 3.9) the most ancestral sequence was from the earliest sampling time point (week 0 / at baseline). But for most of the children (4/9; Study IDs 141586, 131326, 130166, and 134102; Figures 3.7, 3.10, 3.11 and 3.13), the most ancestral sequence was from week 40 of early ART. For 1/9 child (Study ID 146666; Figure 3.6) this sequences was from week 96 of ART and for 2/9 children (Study IDs 147636 and 141806; Figures 3.8 and 3.12), this sequence was from the ART-free period at week 164.

3.2.1.5 NNRTI resistant viruses that replicate during PI-based ART

The NVP-selected resistance mutations Y181C and K101E were the only mutations found between week 40 of ART and the ancestral baseline sequences. Y181C was detected at baseline and week 40 of ART for Study ID 134102. K101E was detected at these two time points for Study ID 147636. The ML trees for these two children show the NVP resistance sequences from baseline and week 40 of ART clustered on the same branch of their respective ML trees. Figure 3.14 is the ML tree for Study ID 134102 and Figure 3.15 is the ML tree for Study ID 147636.

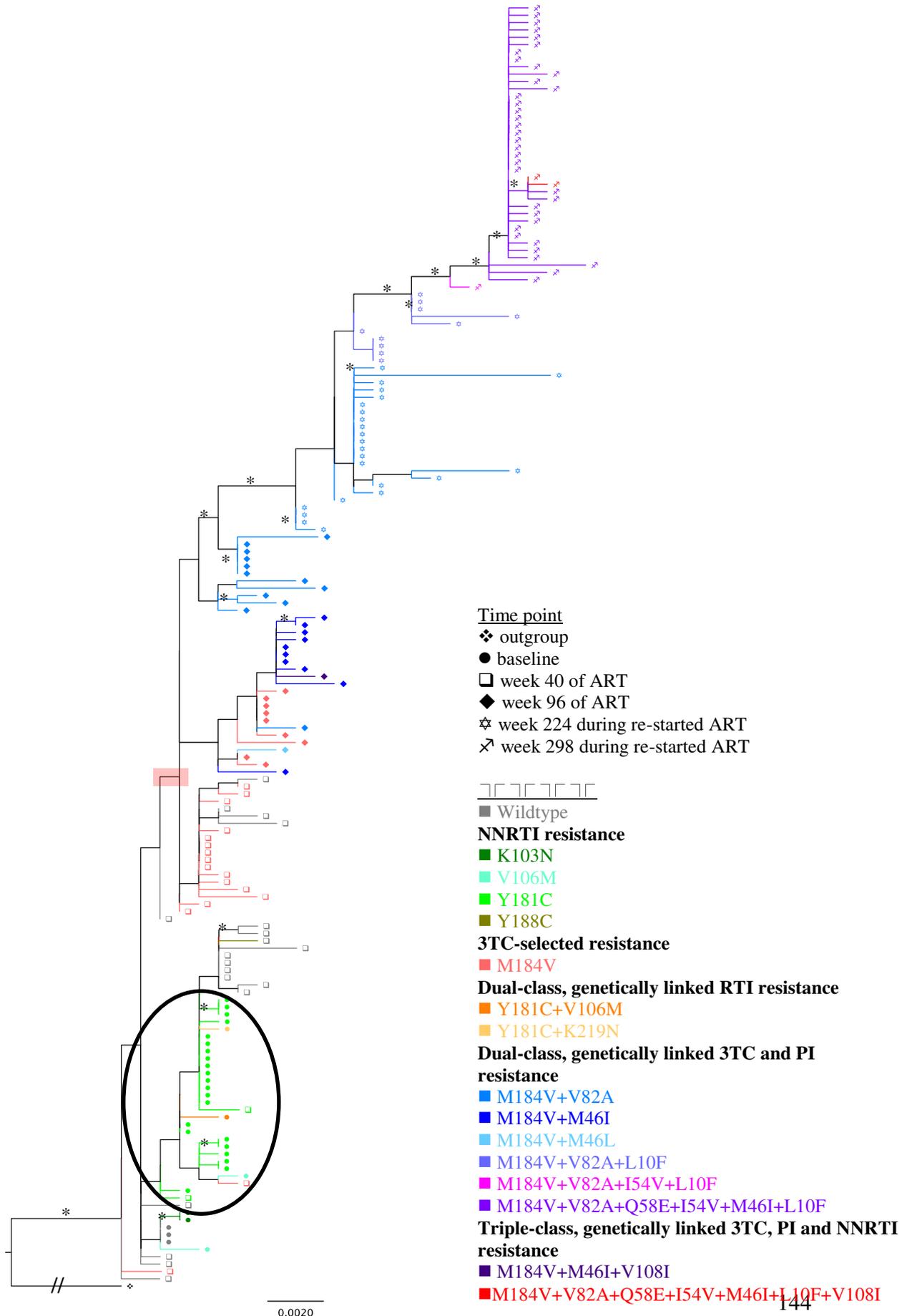


Figure 3.14 ML tree from viral sequences sampled from Study ID 134102. The scale bar represents the number of nucleotide substitutions per site along the branches of the tree. The tree was rooted against an outgroup: an HIV-1 subtype C single genome derived PR-RT sequence from the test sample used to troubleshoot the SGS protocol. * indicated bootstrap values >50. Encircled is the cluster of sequences with NNRTI resistance that persisted from baseline to week 40 of ART. Shaded in red is node at which and M→184→□ was positively selected to give rise to major drug resistant populations.

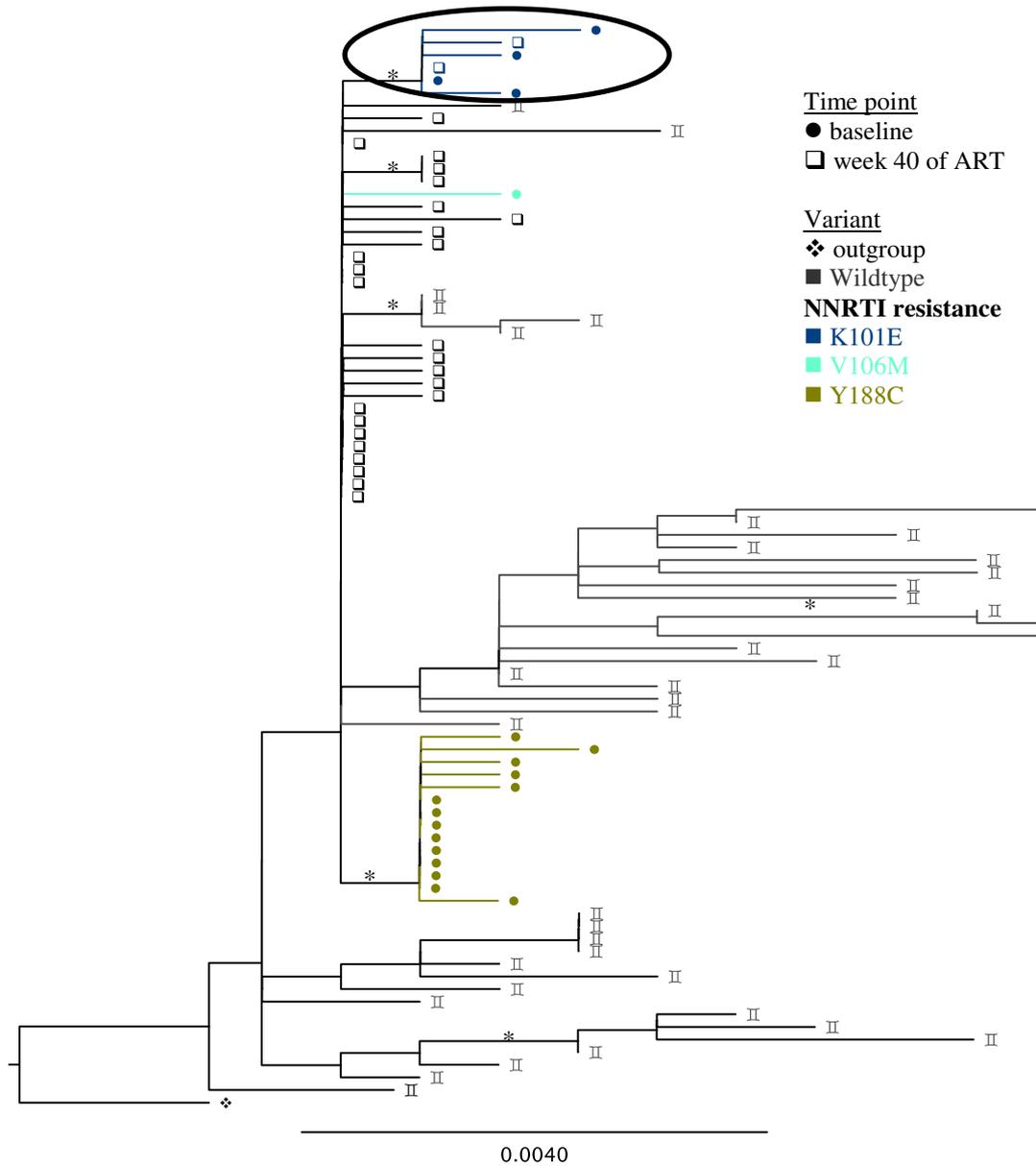


Figure 3.15 ML tree from viral sequences sampled from Study ID 147636. The scale bar represents the number of nucleotide substitutions per site along the branches of the tree. The tree was rooted against an outgroup: an HIV-1 subtype C single genome derived PR-RT sequence from the test sample used to troubleshoot the SGS protocol. * indicated bootstrap values >50. Encircled is the cluster of sequences with NNRTI resistance that persisted from baseline to week 40 of ART.

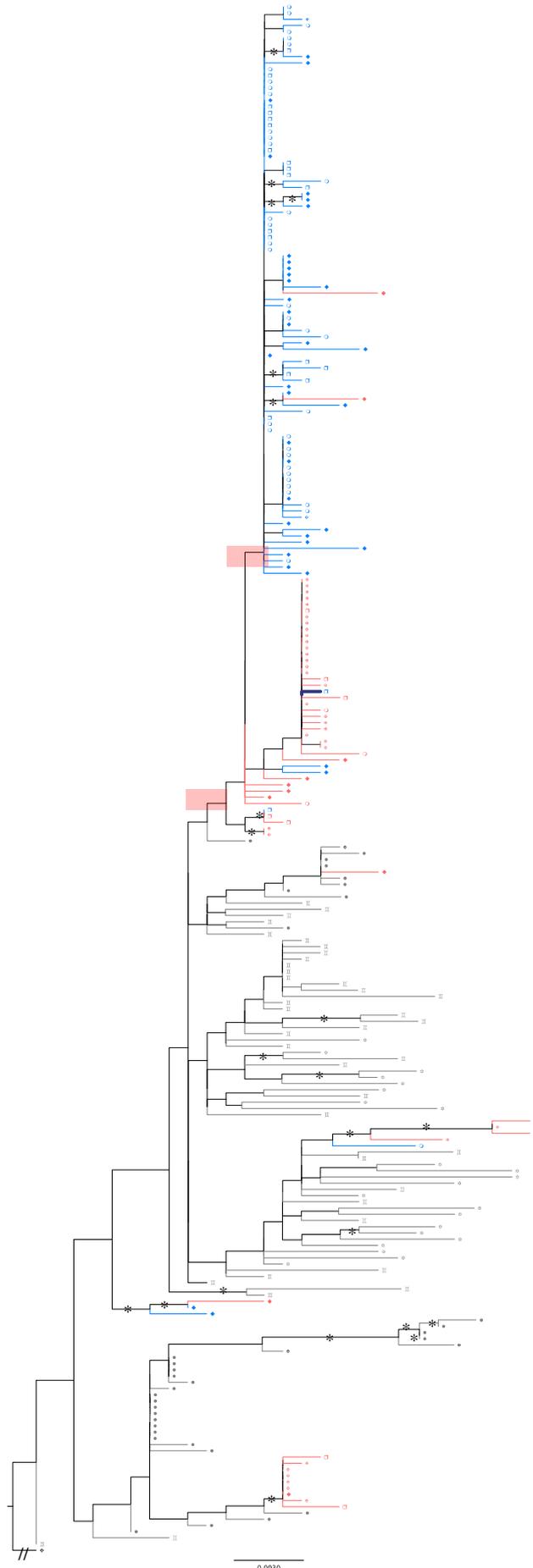


Figure 3.16 ML tree from viral sequences sampled from Study ID 134102. The scale bar represents the number of nucleotide substitutions per site along the branches of the tree. The tree was rooted against an outgroup: an HIV-1 subtype C single genome derived PR-RT sequence from the test sample used to troubleshoot the SGS protocol. * indicated bootstrap values >50. Red highlights are the nodes at which V→82→A and M→184→V were positively selected to give rise to major drug resistant populations.

- ❖ outgroup
- baseline
- ✱ week 12 of ART
- ◻ week 40 of ART
- week 72 of ART
- ◆ week 96 of ART
- ▬ week 164 during ART-free period
- ☆ week 224 during re-started ART

- Variants**
- Wildtype
- NNRTI resistance**
- M184V
- Dual-class, genetically linked 3TC and PI resistance**
- M184V+V82A

3.2.1.6 No evidence of recombination for Study ID 134102

M184V, V82A, M46I and V108I were present in variants detected in the viral population of Study ID 134102 at week 96 of ART, but they were not all linked on the same genome. At week 298, M184V, V82A and M46I were genetically linked in 100% of the viral population and V108I was genetically linked to these mutations in a minority species. However recombination analyses (SBP and GARD) did not reveal evidence of recombination among the sequences obtained from this child, ($p < 0.05$ statistical level), suggesting that multiple linked multi-class drug resistance mutations were not acquired through the recombination of single-class resistant variants detected at week 298. I also did not find any evidence of recombination among sequences from Study ID 141806 with the same methods.

3.2.1.7 Positive selection and co-evolution analyses

FUBAR and FEL analyses for Study ID 141806 revealed 2 positively selected sites in PR-RT common to both analyses (Table 17A). This was the protease inhibitor resistance mutation $V \rightarrow 82 \rightarrow A$ and R38K in RT in these directions of evolution (Table 17A). However these sites did not co-evolve according to Spidermonkey analyses of coevolution. SLAC and FEL analyses for Study ID 134102 did not reveal any positively selected sites but FUBAR analysis revealed 1 positively selected site: $M \rightarrow 184 \rightarrow V$ in RT in this directions of evolution (Table 17B). These nodes are highlighted in Figure 3.14 for Study ID 134102 and Figure 3.16 for Study ID 141806. The mean dN/dS ratios determined by SLAC for both children were < 1 (Table 17) indicating that PR-RT evolution in both children was mainly driven by negative selection.

(A)

Algorithm	Positively selected sites	dN	dS	dN-dS	Number of negatively selected sites	Mean dN/dS ratio
SLAC					28	0.184
FEL	V82A in PR R38K in RT	3.25 4.045	0 5.35e-15	3.25 4.04	54	
FUBAR	R38K in RT V82A in PR D35E in RT R38K in RT N248D in RT	4.045 1.255 1.292 2.007 1.8	2.653e-15 0.191 0.213 0.231 0.243	4.05 1.06 1.08 1.78 1.56	56	

(B)

Algorithm	Positively selected sites	dN	dS	dN-dS	Number of negatively selected sites	Mean dN/dS ratio
SLAC					6	0.375
FEL					15	
FUBAR	M184V in RT	2.352	0.256	9.186	14	

Table 17. Positive selection analysis results for (A) Study ID 141806 and (B) Study ID 134102. SLAC and FEL significance level p-value = 0.05. FUBAR posterior probability = 0.95

3.2.2 Discussion

The most ancestral PR-RT sequence in the ML tree that was re-constructed for most children was not from the baseline population but from a later time point. This is a novel observation to be made in HIV-1 infected children on ART and implies that early ART did not effectively target or eliminate the viral reservoirs. Kieffer et al [239] made a similar observation for the evolution of the *pol* gene in HIV-1 positive adult patients, except that their patients maintained low-level viraemia (viral load <50 copies/ml) during ART and maintained viral loads below the limit of detection without the detection of DRMs for up to 15 months.

I also saw intermingling of viruses from early and late time points despite the development of drug resistance during ART for the majority of these children. For the children who did not develop drug resistance conferring mutations during ART, this result suggests the inhibition of viral evolution in children who received ART. I cannot be conclusive about this statement without comparing these results to those of children who did not receive early ART, but such a group did not exist in the CHER study and early ART became the standard of care at a median follow-up time of 40 weeks (IQR 24–58); 20 weeks (IQR 16–25) for the deferred therapy group, because early ART was found to reduce the risk of death by 75% during the trial.

The strong selection pressure exerted by ART was responsible for the inhibition of viral evolution from weeks 12 to 96 of ART for Study ID 141806, which is why the majority of drug resistant sequences adhered to the backbone of the ML tree. In contrast there was continuous evolution of the drug resistant population of Study ID

134102 from baseline and until week 298 during re-started ART with the accumulation of drug resistance mutations on the viral genomes. However similar to the findings of Shankarappa et al [240], ART seemed to inhibit viral replication more efficiently during re-started ART when the sequences from weeks 224 and week 298 were more similar than within the previous time points and were the most similar, i.e. adhered to the backbone of the tree, at week 298. MTCT-NVP seemed responsible for this inhibition in the baseline population since this child did not receive prophylactic NVP.

Recently Kearney et al [222] proposed that “the presence of identical sequences from difference sampling time points before and during ART suggested that there was a proliferating infected cell population that was the main source of viraemia”.

Intermingled identical sequences from baseline and week 40 of ART, suggested that such an infected cell population was the source of persisting viraemia in 4 children until week 40 (2 children who received ART until week 40 and 2 children who received continuous ART). Kearney et al [222] also suggest that the viral reservoir may “contract during prolonged therapy”. Particularly in the case of Study ID 143646, my data suggests that the combination of prophylaxis, early ART and continuous therapy may have the potential to restrict the viral reservoir in the first instance and then reduce it (e.g. Figure 3.9). It was very unlikely that there was cross-contamination between the samples from these two time points because these samples were received and processed independently (~12 months apart).

The phylogeny for Study ID 141806, shows M184V single-mutated sequences as selected independently on multiple occasions, but only one of these events gave rise

to the major contributor to drug resistance during ART. Since the major M184V single-mutated population shared a common ancestor with the major double-mutated M184V+V82A population and V82A single-mutated sequences were never detected. I made the same observation for single-mutated M184V populations and double-mutated M184V+V82A populations based on the phylogeny for Study ID 134102. V82A single-mutants were also not detected at any of the sampling time points for this child.

Similar to my findings, M184V is known to be detected within 12 weeks of a 3TC inclusive ART regime [241, 242] and in combination with my phylogenetic analyses, I propose that V82A was only selected in M184V-mutated viruses. This finding may speak to the action of combination therapy suggesting that only when the action of 3TC is inhibited at the start of the viral replication cycle by the M184V mutation, does the action of LPV at the end of the replication cycle become relevant enough to cause selection and proliferation of V82A in the viral population. There is still disagreement about the effect of NRTIs on cell-to-cell spread of HIV-1 [243, 244], but based on the findings of Sigal et al [244], that NRTIs are not very effective at blocking cell-to-cell spread, I hypothesize that when 3TC is no longer inhibiting viral replication through chain termination, the second NRTI, AZT is even less likely to inhibit cell-to-cell spread of M184V mutated viruses. Consequentially, this would have allowed the PI (LPV) to exert an even stronger selection pressure, resulting in M184V and V82A double-mutated viruses.

These findings also highlight the fact that drug resistance can persist despite efficient suppression of viral replication by ART and HIV-1 can escape treatment recycling

through the proliferation of multiple linked multi-class drug resistant viruses from such a reservoir. In contrast, drug resistance mutations were not detected during treatment interruption (week 164) or during re-started ART (week 224) and by week 265 of re-started ART for Study ID 141806. Instead, viral replication was suppressed to undetectable levels (<400 copies/ml at weeks 265 and 281), possibly because rebounding viraemia after treatment interruption was a result of archived wildtype viruses that were susceptible to re-cycled ART.

My data suggests that NVP-resistant viruses detected at week 40 of ART were proliferating from the NVP-resistant reservoir that was producing these viruses at baseline. Many studies have concluded that failed prophylactic NVP is associated with failure of NNRTI-based ART in HIV-1 infected adults and children and an obvious deduction from this is that an NNRTI-resistance viral reservoir can be established from NNRTI-resistant viruses that were selected by prophylactic NVP. To my knowledge, my study demonstrated the persistence of these viruses despite the lack of other published data showing that these viruses can persist in the viral population during PI-based ART.

M184V in RT and V82A in PR did not co-evolve despite evidence that V82A was only selected on a background of M184V mutated sequences. Without this phylogenetic analysis, I might have assumed an evolutionary relationship between these two sites in PR and RT. But my phylogenetic data implied that the development of V82A in these variants was not dependent on the development of M184V in RT. Therefore I considered a lack of sensitivity of the methods used to detect co-evolution between M184V and V82A, but to my knowledge, Spidermonkey [221] is currently

the most rigorous co-evolution analysis platform because it considers evolutionary history of my data, the phylogenetic relationships of my intra-host sequences and utilizes coupled nucleotide substitution models in order to avoid predictions of false positives [245-247].

When *Vif* does not efficiently sequester APOBEC3G/F, it can cause guanine-to-adenine transitions in the viral genome, which can lead to premature STOP codons in the viral genome and sometimes point mutations that can give rise to drug resistance mutations. These APOBEC-induced mutations usually exist at a low frequency in the viral population of adults and children, but since SGS is able to detect low-level variants in the viral population, these mutated sequences can confound analyses aimed at determining ART-selected drug resistance, population diversity and phylogenetic reconstructions [248] [180]. Since there is no known correlation between disease progression and frequency of APOBEC-induced mutations in the HIV-1 genome in infected children [249] I analysed the single genome sequences from all the sampling time points in each child (Study IDs 141806 and 134102) in a single MSA per child.

There are very few studies on the prevalence or contribution of APOBEC3G/F-induced guanine-to-adenine transitions contribute to the diversity of HIV-1 *pol* in infected children, and I found only one that addressed the contribution of this cytidine deamination of *pol* in children [250]. They determined that HIV-1 *pol* was one of the most affected genes in vertically infected children [250] but it did not surpass a prevalence of 13%. Therefore I did not expect or observe APOBEC3G/F cytidine deamination in *pol* to surpass 13% compared to the most ancestral sequence obtained per child.

3.2.3 Conclusions

In my study, when M184V is detected with PI resistance and as genetically linked on the same viral genome, my evidence suggests that PI-resistance mutations are only selected in M184V mutated sequences selected by 3TC during PI-based ART. But the evolution of PI resistance mutations in PR and M184V in 3TC were not shown to be dependent on each other. NNRTI resistance detected by SGS at baseline and week 40 of PI-based ART are examples of persistence of PMTCT-NVP selected resistance from a reservoir of NNRTI resistant virus. Further research, such as prospective studies with larger cohorts is needed to generalise these findings for children who receive PI-based ART.

I also saw that early ART did not effectively eliminate the viral reservoir so that it continued to be a major contributor to the evolution of the viral populations during virological success and failure on ART; highlighting the importance of the viral reservoir. I also suggest that prophylaxis and early ART may not just be strategies to improve child mortality and disease progression, but may also be strategies to restrict the viral reservoir in the first instance and then reduce it. This aspect of early ART needs to be researched further.

My findings do not give more clarity to the effectiveness of recycled PI-based ART regimes in children because I saw the re-emergence and susceptibility of viruses during re-cycled ART in one of the children who developed majority species drug resistance during the first round of ART, while multiple linked multi-class drug resistance persisted in the other child who continued to fail a recycled regime of re-started ART.

3.3 Longitudinal trends in replication capacity and drug susceptibility associated with development of multiple linked multi-class drug resistance and characteristics of Gag during PI-based ART

In this chapter I examine the drug susceptibility and replication capacity (RC) conferred by multi-class drug resistant PR-RT from two children who developed such resistance in their majority species during ART (Study ID 141806 and 134102). In previous chapters I described the genotypic development of variants in these two children and the persistence of drug resistance with time.

PR-RT from selected single genome derived PR-RT sequences were cloned into the gag-pol expression vector p8MJ4-HSC to produce a chimeric *gag-pol* expression vector. I used an MSA of the cloned patient PR-RT amino acid sequences to characterize amino acids that were known to affect PI and RTI susceptibility and amino acids that changed over time. I used single replication cycle assays to determine replication capacity (RC) and drug susceptibility of pseudoviruses produced from these chimeric vectors.

The RC of pseudoviruses produced from each chimeric vector was determined from titrations of serially diluted viruses on 293T cells for 48 hours. Infection was determined by measuring the luciferase expression of infected target cells. An enzyme linked immunosorbent assay was used to measure the reverse transcriptase activity of each pseudovirus. RC was measured as RLU/ng of RT activity and expressed as % of the WT for each child.

I measured the drug susceptibility of the pseudoviruses as the concentration of the drug that inhibited pseudoviral replication by 50% (IC₅₀) and susceptibilities were

expressed as fold change in IC₅₀ relative to the baseline WT pseudovirus produced for each child.

I tested the susceptibility of these viruses to the components of early ART (AZT, 3TC and LPV) as well as the components of the recommended second line therapy for these children (Abacavir (ABC), Didanosine (ddI), Efavirenz (EFV) and NVP). I also tested their susceptibilities to two other PIs that have been approved for use in children, namely Nelfinavir (NFV) and Saquinavir (SQV).

I also tested DRV susceptibility of the pseudovirus that was singly mutated with M184V and the pseudovirus from week 298 of ART from Study ID 134102 which had multiple major and minor PI drug resistance mutations in PR (V82A, I54V, Q58E, M46I and L10F). The fold changes in IC₅₀ of these two viruses were measured relative to the HIV-1 subtype C gag-pol expression vector p8MJ4-HSC. Each drug mentioned in this chapter, along with its abbreviation and drug class are listed in Appendix C Table 24.

Finally I used population sequence analysis of HIV-1 Gag from the sampling time points of Study IDs 141806 and 134102 to determine the involvement of Gag that may have also affected PI susceptibility. I identified amino acids in Gag that have been associated with PI susceptibility in the literature and amino acid changes over time at positions associated with PI-exposure. A position specific scoring matrix (PSSM) was produced from an MSA of the 891 HIV-1 Subtype C Gag sequences from treatment naïve children from Sub-Saharan African children that were available in the HIV Los Alamos sequence database on June 1st 2014. I used the PSSM to determine the natural variation of amino acids found at these key positions in Gag for my study cohort using population sequence analysis.

3.3.1 Results

3.3.1.1 Longitudinal amino acid changes in PR-RT

I replaced PR-RT in the p8MJ4-HSC plasmid with PR-RT from select single genome sequences. Obtained were 4 chimeric vectors for Study ID 141806 and 6 chimeric vectors for Study 134102 that produced pseudovirus based on retroviral titrations. The amino acids in PR and RT associated with drug susceptibility or replication capacity of HIV-1 that defined each chimeric vector are listed in Table 17 for Study ID 141806 and Table 18 for Study ID 134102.

	Pseudovirus	1	2	3	4	Background amino acids associated with RC and PI-exposures/resistance
Viral element	Amino acid position	Amino acid changes				
<i>pol</i> TFP	50	V	A	A	A	
Protease	14	R	K	K	K	M36I, R41K, L63T, I93L
	20	K	R	R	R	
	35	E	D	D	D	
	82	V	V	A	V	
	94	G	A	A	G	
Reverse transcriptase	36	D	E	E	D	
	40	E	D	D	E	
	184	M	V	V	M	
	250	N	D	D	N	

Table 18. Amino acids in PR-RT associated with RC or drug resistance for Study ID 141806. Longitudinal amino acid changes in *pol* for Study ID 141806 in columns 3 to 6. Background amino acids present in all pseudoviral PR-RT sequences are listed in column 7. Identifying elements of pseudoviruses are: (1) WT from baseline (2) 184V single-mutated sequence from week 12 of early ART (3) 184V and 82A double-mutated sequence representing the majority consensus from weeks 40, 72 and 96 of early ART (4) Sequence from week 164 without known DRMs and during ART interruption

	Variant	1	2	3	4	5	6	Background amino acids associated with RC and PI-exposures/resistance
Viral element	Amino acid position	Amino acid						
<i>pol</i> TFP	9	P	P	P	P	L	L	K20R, R41K, L63P, V77I, L89M and I93L
Protease	10	L	L	L	L	F	F	
	46	M	M	M	M	I	I	
	54	I	I	I	I	V	V	
	58	Q	Q	Q	Q	E	E	
	82	V	V	V	A	A	A	
98	N	N	S	N	N	N		
Reverse transcriptase	108	V	V	V	V	V		I
	162	C	C	C	S	S		S
	174	Q	Q	Q	Q	R		R
	181	Y	C	C	Y	Y		Y
	184	M	M	V	V	V		V
	277	R	R	R	R	K		K
	292	V	V	I	I	I		V
	296	N	N	T	T	T		N
312	E	E	Q	Q	Q	E		

Table 19. Amino acids in PR-RT associated with RC or drug resistance for Study ID 134102. Longitudinal amino acid changes in *pol* for Study ID 134102 in columns 3 to 8. Background amino acids present in all pseudoviral PR-RT sequences are listed in column 7. TFP is the transframe protein. Identifying elements of pseudoviruses are: (1) WT from baseline (2) 181C single-mutated sequence from week 40 of early ART (3) 181C single-mutated sequence from week 96 of ART (4) 184V and 82A double-mutated sequence from week 224 during re-started ART (5) 184V and 82A, 54V, 58E, 46I and 10F multiple-mutated sequence from week 298 during re-started ART (6) 108I, 184V, 82A, 54V, 58E, 46I and 10F multiple-mutated sequence from week 298 during re-started ART.

3.3.1.2 Single round infectivity of pseudoviruses

To determine the phenotypic impact of DRMs that accumulate in PR-RT. I used the TCID₅₀ and RT-activity (ng of cDNA produced) of each pseudovirus to measure RC as RLU/ng of RT-activity as a % of WT (Appendix C Figure 5.2A and 5.2B).

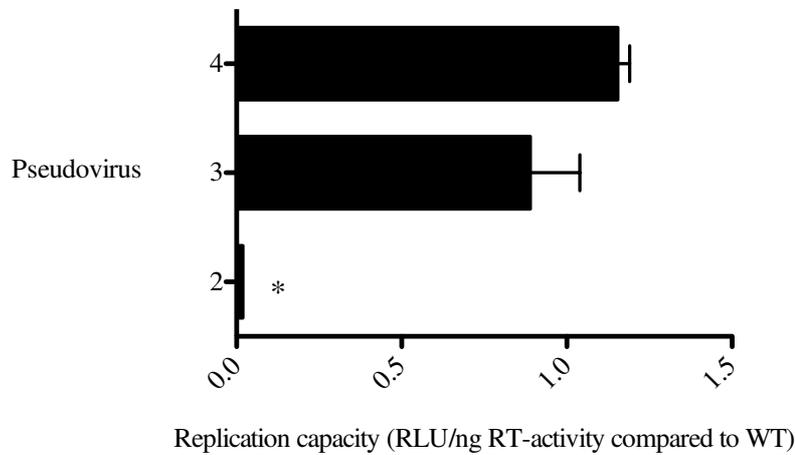
When M184V in RT was the only DRM present, the single round infectivity (as a surrogate for) replication capacity of these pseudoviruses was less than the WT (RC was 2% of the WT for Pseudovirus 2 from Study ID 141806 and 50% of the WT for Pseudovirus 3 from Study ID 134102).

For Study ID 134102, the RC of the Y181C mutated pseudovirus (Pseudovirus 2) was 80% less than the WT. The variance of RC for Pseudovirus 3 did not conclusively determine if the RC recovered with the PR mutation N98S and the RT mutations: M184V, V292I, N296T and E312Q. However Pseudovirus' 4, 5 and 6 had comparable RCs to WT with the mutations seen in Pseudovirus 3 as well as additional mutations in PR and RT: Pseudovirus 4 had the same mutation profile as Pseudovirus 3 plus V82A in PR and M184V, C162S in RT. Pseudovirus 5 also had the same mutations profile as Pseudovirus 4 as well as L10F, M36I, I54V, Q58E in PR and Q174R in RT. Pseudovirus 6 had the same mutation profile as Pseudovirus 5 plus V108I in RT. K20R in PR is associated with increased fitness of HIV-1 when detected with V82A in PR [102] and is present as a background mutation in all of these pseudoviruses.

For Study ID 141806, Pseudovirus 2 had the mutations R14K, K20R, E35D in PR and D36E, E40D, M184V, N250D in RT and had an RC that was 80% less than the WT. Pseudoviruses 3 and 4 ha RCs that were comparable to WT with additional

mutations. Pseudovirus 3 had the mutation profile of Pseudovirus 2 plus V82A in PR. In Pseudovirus 4, G94A, V82A in PR and D36E, E40D, M184V and N250D reverted to the WT amino acids at these positions. K20R was not present in the WT pseudovirus, but was present in the other 3 pseudoviruses (Pseudoviruses 2, 3 and 4).

(A)



(B)

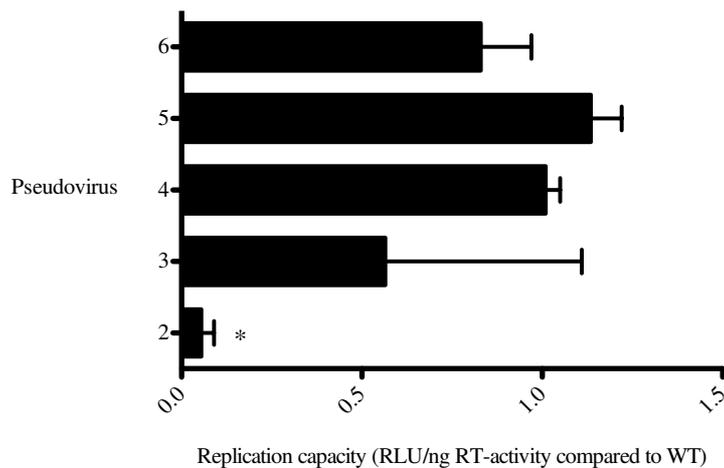


Figure 3.17 Replication capacity expressed as relative light units (RLU) per ng of RT-activity as % of the RC of WT pseudovirus. Therefore when the mean RC = 1, this means equal RC compared to the WT. * above the bar of the RC indicated that the mean RC of 2 independent experiments was significantly different from that of Pseudovirus 1 (Student's t-test; $p < 0.05$). (A) Relative RC for the pseudovirus panel of Study ID 141806 (B) Relative RC for the pseudoviral panel of Study ID 134102.

3.3.1.3 Drug susceptibility

The phenotypic susceptibility of viruses from the 2 study patients was assessed as detailed above. The mean IC₅₀ and standard error of the mean IC₅₀ (SEM) were plotted for each child and for each drug class. The baseline WT virus was assigned “Pseudovirus 1” for both children and its IC₅₀ for all the drugs tested in my study was comparable to the original gag-pol expression vector p8MJ4-HSC (Appendix C Figure 5.1A and 5.1B).

2.1.1.1 Study ID 134102

2.1.1.1.1 PI resistance

As expected, pseudoviruses that did not have known PI resistance conferring mutations in PR (Pseudoviruses 2 and 3) had comparable IC₅₀ to WT. Pseudovirus 4 had a 5-fold increase in SQV IC₅₀ ($p = 0.0133$; IC₅₀ = 2.22nM \pm 0.0805) and 8-fold increase in NFV IC₅₀ ($p = 0.0027$; IC₅₀ = 13.91nM \pm 1.36) with V82A and K20R in PR, whereas its LPV IC₅₀ was comparable to WT (2-fold change in IC₅₀, $p = 0.33$; IC₅₀ = 0.3986nM \pm 0.1602). Pseudoviruses 5 and 6 had >200-fold changes in LPV IC₅₀ compared to WT ($p = 0.0419$; IC₅₀ = 69.5nM \pm 14.66 and $p = 0.0059$; IC₅₀ = 54.71nM \pm 4.19), >20-fold changes in NFV IC₅₀ ($p = 0.0002$; IC₅₀ = 105.9nM \pm 1.25 and $p = 0.0113$; IC₅₀ = 40.53nM \pm 14.26) and >10-fold change in SQV IC₅₀ ($p = 0.0007$; IC₅₀ = 5.504nM \pm 0.1295 and $p = 0.0003$; IC₅₀ = 4.801nM \pm 0.0545).

High-level PI-resistance, especially for LPV, was expected of these two pseudoviruses because they had the major PI-DRMs M46I, I54V and V82A [102] as well as the minor PI-DRMs L10F and Q58E, which further enhance LPV resistance and cross-resistance to the other PIs. (L10F is a minor PI mutation that is associated

with reduced susceptibility to all the PIs except SQV, tipranavir (TPV) and atazanavir (ATV) and Q58E is associated with reduced susceptibility to all but TPV [102]. Pseudoviruses 3 and 5 were susceptible to DRV. These results are shown in Figure 3.18.

2.1.1.1.2 RTI resistance

Only pseudoviruses with known NNRTI resistance conferring mutations in RT had higher fold-changes in EFV and NVP IC50s compared to WT. Pseudovirus 2 was Y181C-mutated from week 40 of ART and it had a 209-fold increase in NVP IC50 ($p = 0.01626$; $IC_{50} = 29.04nM \pm 13.4$) and 18-fold increase in EFV IC50 ($p = 0.0111$; $IC_{50} = 11.81nM \pm 1.222$) compared to WT, which is in-keeping with the literature [102].

Pseudovirus 6 was V108I mutated and also had a 35-fold increase in NVP IC50 ($p = 0.0161$; $IC_{50} = 69.4nM \pm 8.665$) but had comparable susceptibility for EFV as WT.

All pseudoviruses with M184V in RT (Pseudoviruses 3-6) had >400-fold increase in 3TC IC50 ($p = 0.0370$; $IC_{50} = 29.62nM \pm 5.789$, $p = 0.0009$; $IC_{50} = 33.02nM \pm 0.957$, $p = 0.0055$; $IC_{50} = 87.5nM \pm 6.48$ and $p = 0.0011$; $IC_{50} = 34.44 \pm 1.109$, respectively). None of the viruses had a significantly different mean AZT IC50, ABC IC50 or ddI IC50 compared to WT. This was expected, given that no known TAMs were present in RT of these pseudoviruses and M184V has only been shown to confer low-level resistance to ABC and potential low-level resistance to ddI [102]. These results are shown in Figure 3.19.

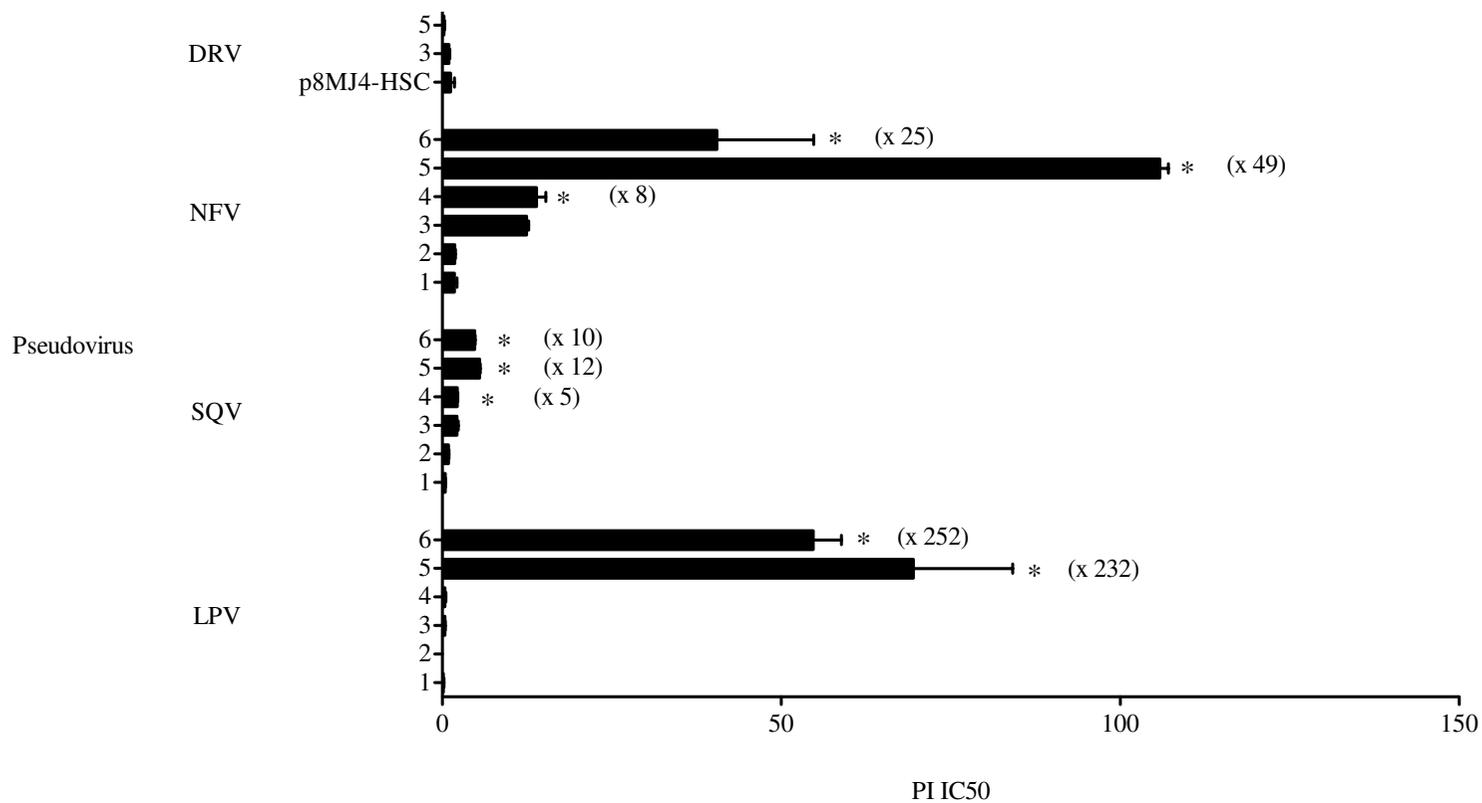


Figure 3.18 PI IC50s of Pseudoviruses 1-6 for Study ID 134102. * above the bar of the PI IC50 indicated that the mean IC50 of 2 independent experiments was significantly different from that of Pseudovirus 1 (Student's t-test; $p < 0.05$). In parentheses is the fold change in PI IC50 compared to Pseudovirus 1. Each PI tested is indicated on the y-axis

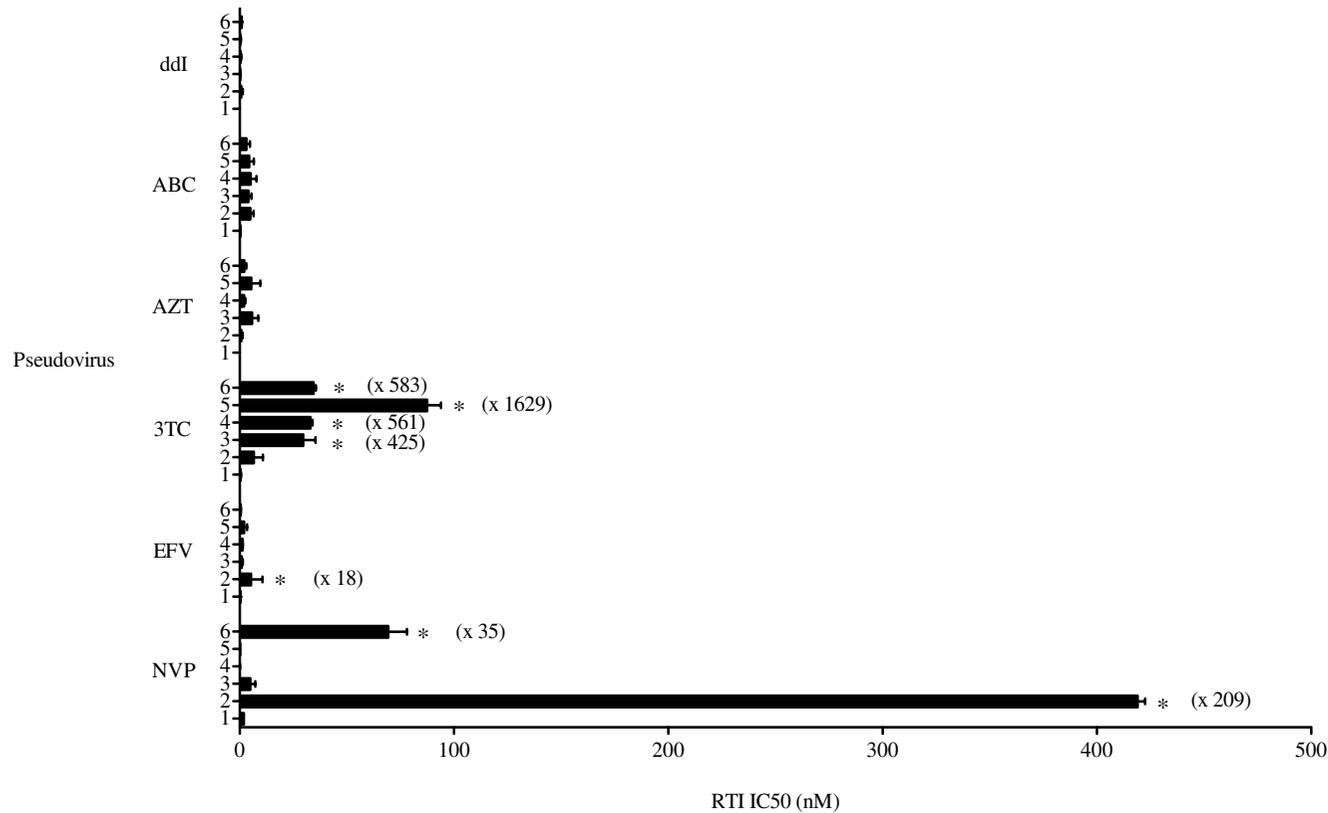


Figure 3.19 RTI IC50s of Pseudoviruses 1-6 for Study ID 134102. * above the bar of the RTI IC50 indicated that the mean IC50 of 2 independent experiments was significantly different from that of Pseudovirus 1 (Student's t-test; p < 0.05). In parentheses is the fold change in RTI IC50 compared to Pseudovirus 1 when the two means were different. Each RTI tested is indicated on the y-axis

2.1.1.2 Study ID 141806

2.1.1.2.1 PI resistance

As expected, Pseudovirus 3 had V82A in PR and had an 11-fold increase in SQV IC₅₀ and 27-fold increase in NFV IC₅₀ compared to WT ($p = 0.0053$; IC₅₀ = $0.9962\text{nM} \pm 0.05385$ and $p = 0.0052$; IC₅₀ = $13.65\text{nM} \pm 0.85$). However Pseudovirus 4, which had no known major PI-resistance conferring mutations in PR, had a 14-fold increase in SQV IC₅₀ and a 25-fold increase in NFV IC₅₀ compared to WT ($p = 0.0052$; IC₅₀ = $2.255\text{nM} \pm 0.148$ and $p = 0.0297$; IC₅₀ = $13.73\text{nM} \pm 2.26$).

Pseudovirus 4 differed from Pseudovirus 1 in PR with the mutations R14K, K20R and E35D. These results are shown in Figure 3.20.

2.1.1.2.2 RTI resistance

None of the pseudoviruses had significantly higher NVP IC₅₀s or EFV IC₅₀s compared to WT, and they did not harbour any known NVP or EFV resistance conferring mutations in RT. Pseudoviruses 2 and 3, both had M184V in RT and 148- and 158-fold increase in 3TC IC₅₀ ($p = 0.0388$; IC₅₀ = $16.77\text{nM} \pm 3.38$ and $p = 0.0416$; IC₅₀ = $19.33\text{nM} \pm 4.046$). All but Pseudovirus 3 were susceptible to AZT, ddI and ABC. Pseudovirus 3 had 84-fold increase in ABC IC₅₀ ($p = 0.027$; IC₅₀ = $6.692\text{nM} \pm 1.108$) and 25.5-fold increase in ddI IC₅₀ ($p = 0.0399$; IC₅₀ = $0.4871\text{nM} \pm 0.0635$). These results are shown in Figure 3.21.

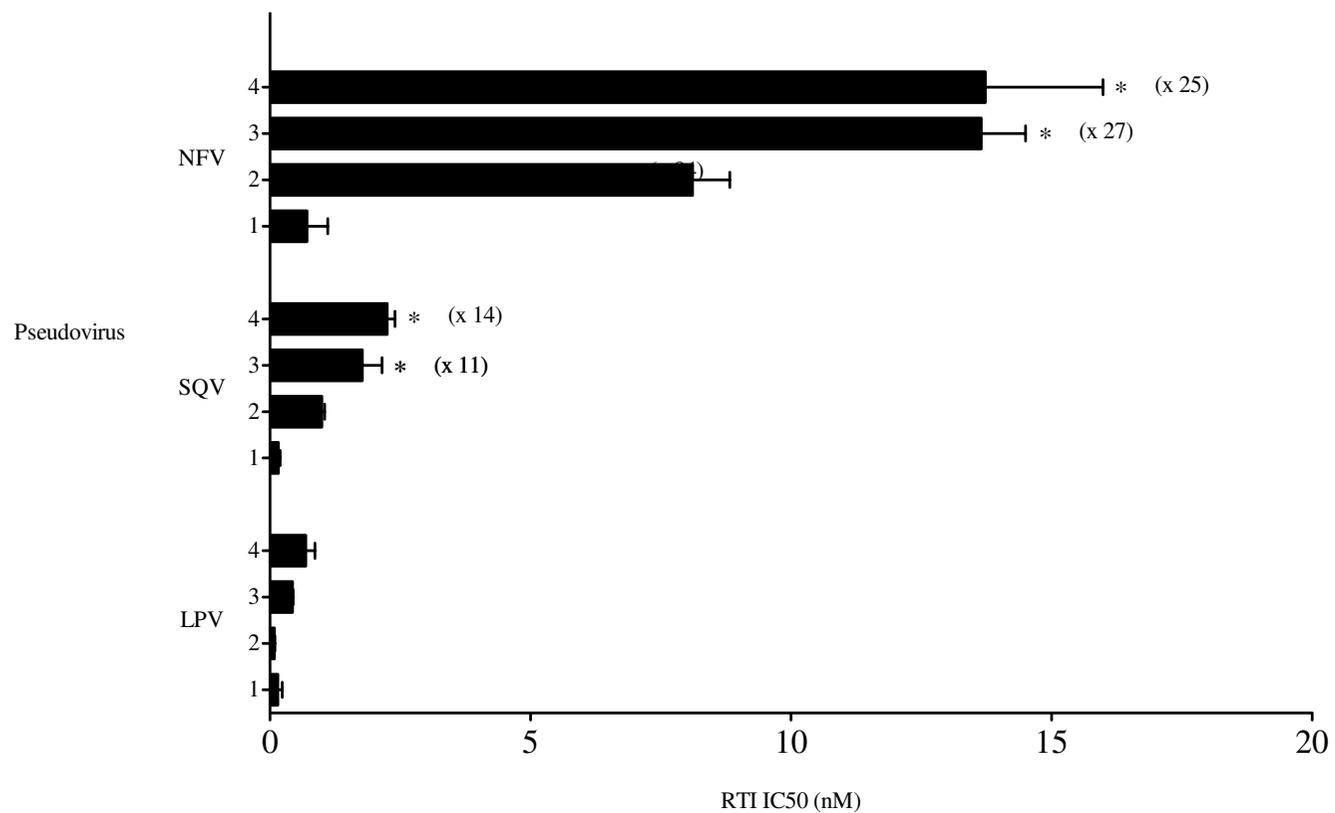


Figure 3.20 PI IC50s of Pseudoviruses 1-4 for Study ID 141806. * above the bar of the PI IC50 indicated that the mean IC50 of 2 independent experiments was significantly different from that of Pseudovirus 1 (Student's t-test; $p < 0.05$). In parentheses is the fold change in PI IC50 compared to Pseudovirus 1 when the two means were statistically different. Each PI tested is indicated on the y-axis.

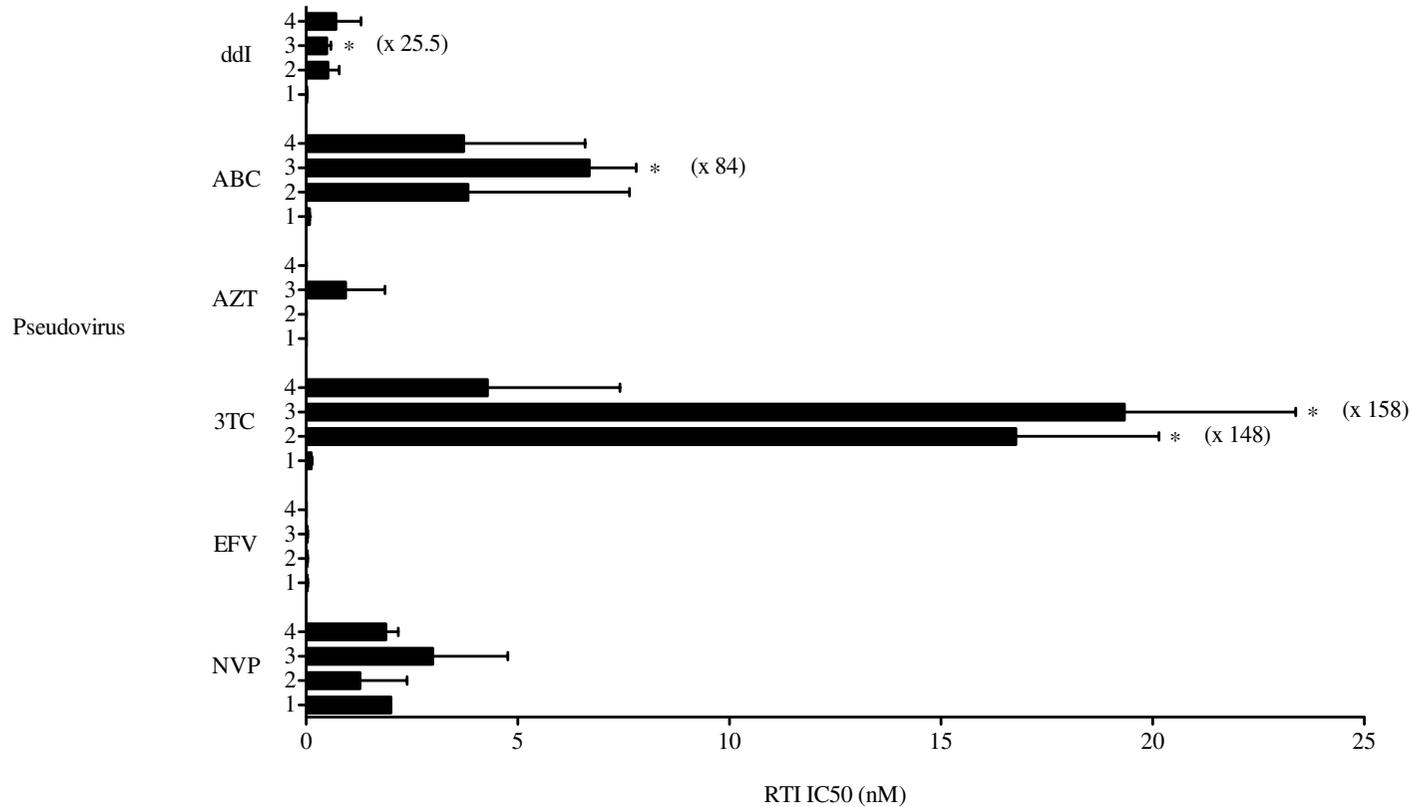


Figure 3.21 RTI IC50s of Pseudoviruses 1-4 for Study ID 141806. * above the bar of the RTI IC50 indicated that the mean IC50 of 2 independent experiments was significantly different from that of Pseudovirus 1 (Student's t-test; $p < 0.05$). In parentheses is the fold change in RTI IC50 compared to Pseudovirus 1 when the two means were different. Each RTI tested is indicated on the y-axis.

3.3.1.4 Variation of Gag

I analysed the population sequence of gag-PR-RT at baseline for the 10 children in my study cohort and at week 40 of ART for 8 children. To determine changes in Gag with the development of PI-resistance mutations in PR, I also analysed the population sequence of Gag at weeks 0, 12, 40, 72, 96 and 164 for Study ID 141806 and at weeks 0, 40, 96, 224 and 298 for Study ID 134102 (Tables 20 and 21).

3.3.1.4.1 Variation of Gag cleavage sites

PSSM analysis of 891 HIV-1 subtype C gag sequences derived from treatment naïve children from Sub-Saharan Africa (HIV Los Alamos Sequence database) revealed high conservation ($\geq 90\%$) of the amino acid sequence of the p17/p24 and p24/p2 cleavage sites of HIV-1 subtype C sequences in children compared to HXB2 and subtype B sequences [251]. The amino acid sequences of these two cleavage sites from children within my study cohort were also identical to subtype B and C consensus sequences (Table 20).

The amino acid sequence of the p2/p7 cleavage site remained identical between week 0 (baseline) and 40 of ART of the study cohort (n=10) and at each sampling time point for Study ID 141806. Asparagine (N) has a frequency of 70% and serine (S) has a frequency of 22% at position 375 in the p2/p7 cleavage site of paediatric treatment naïve sequences. For Study ID 134102, (N) was replaced with (S) at position 375 at week 40 of ART, but reversion to N was detected at week 96 and subsequent sampling time points. However at week 96 of ART, methionine (M) was replaced with lysine (L) at position 378 and M378L was maintained at the subsequent

sampling time points for this child. This coincided with glutamic acid (E) being replaced by aspartic acid (D) at position 428 in the p7/p1 cleavage site, which was also maintained at subsequent sampling time points for this child. 378M and 428E are highly conserved in treatment naïve paediatric subtype C sequences (both have frequency of 97%), whereas 378L and 428D are rare at these positions (0.6% and 0.1% respectively). The p7/p1 cleavage site was conserved between time points for Study ID 141806 and between week 0 and week 40 of ART for the study cohort.

The p1/p6 cleavage site was conserved between sampling time points for Study IDs 141806 but for Study ID 134102, proline (P) was replaced with lysine (L) at position 453 at week 298 during re-started ART, where L is rarely seen at this position (frequency = 7% in subtype C infected treatment naïve children) and P is the most frequently observed residue (frequency = 86%). N451S was detected in the p1/p6 cleavage site between week 0 and week 40 of ART in the study cohort and both residues have comparable frequencies within this cleavage site of treatment naïve subtype C infected children (N = 54.7% and S = 42.4%).

3.3.1.4.2 Variation of structural elements of Gag, p2 and p6

In addition to the cleavage site mutations described in the previous section, I also detected mutations in structural elements of Gag, p2 and p6 (Table 21) that were associated with PI exposure as detailed in the literature [154]. All sequences from the consensus study cohort sequences from baseline and week 40 of ART as well as Study IDs 141806 and 134102 had an amino acid deletion at position 371 in p2. This was not a surprise given that my PSSM of subtype C sequences revealed a 98% frequency of a deletion at this position in Gag. All sequences obtained from Study ID

141806 had Y79F in p17 where the frequency of phenylalanine (F) at this position was 48.5% and the frequency of tryptophan (Y) was 43% based in my PSSM result. T487I in p6 was detected at week 12 of ART (the sampling time point before V82A was first detected in PR, i.e. week 40) and continued to be detected at all subsequent time points. Isoleucine (I) (22%) was not as common as threonine (T) (77%) at position 487 of Gag based on the PSSM. Their detection also coincided with the emergence of the minor PI-resistance mutation, K20R in PR.

For Study ID 134102, V82A in PR was first detected in this child at week 96 of ART. In Gag, I401L in p7 and E428D in the p7/p1 cleavage site were also first detected at this time point and at all subsequent time points. I (82%) was the most common amino acid at position 401 and E (97%) was the most common amino acid at position 428 based on the PSSM).

		Protease										Gag cleavage sites				
Amino acid position		10	14	20	35	46	54	58	82	94	98	p17/p24	p24/p2	p2/p7	p7/p1	p1/p6
											128-137	359-368	373-382	428-437	444-453	
HXB2		L	R	K	E	M	I	Q	V	G	N	VSQNY/PIVQN	KARVL/AEAMS	SATIM/MQRGN	ERQAN/FLGKI	RPGNF/LQSRP
134102	Baseline	L	R	R	E	M	I	Q	V	G	N	VSQNY/PIVQN	KARVL/AEAMS	NINIM/MQKSN	ERQAN/FLGKI	RPGNF/LQNRP
	Week 40	L	R	R	E	M	I	Q	V	G	N	VSQNY/PIVQN	KARVL/AEAMS	NISIM/MQKSN	ERQAN/FLGKI	RPGNF/LQNRP
	Week 96	L	R	R	E	M	I	Q	A	G	S	VSQNY/PIVQN	KARVL/AEAMS	NINIM/LQKSN	DRQAN/FLGKI	RPGNF/LQNRP
	Week 224	L	R	R	E	M	I	Q	A	G	N	VSQNY/PIVQN	KARVL/AEAMS	NINIM/LQKSN	DRQAN/FLGKI	RPGNF/LQNRP
	Week 298	F	R	R	E	I	V	E	A	G	N	VSQNY/PIVQN	KARVL/AEAMS	NINIM/LQKSN	DRQAN/FLGKI	RPGNF/LQNR L
141806	Baseline	L	R	K	E	M	I	Q	V	G	N	VSQNY/PIVQN	KARVL/AEAMS	NANIM/MQRGN	ERQAN/FLGKI	RPGNF/LQNRP
	Week 12	L	K	R	E	M	I	Q	V	G	N	VSQNY/PIVQN	KARVL/AEAMS	NANIM/MQRGN	ERQAN/FLGKI	RPGNF/LQNRP
	Week 40	L	K	R	D	M	I	Q	A	G	N	VSQNY/PIVQN	KARVL/AEAMS	NANIM/MQRGN	ERQAN/FLGKI	RPGNF/LQNRP
	Week 72	L	K	R	D	M	I	Q	A	G	N	VSQNY/PIVQN	KARVL/AEAMS	NANIM/MQRGN	ERQAN/FLGKI	RPGNF/LQNRP
	Week 96	L	K	R	D	M	I	Q	A	G	N	VSQNY/PIVQN	KARVL/AEAMS	NANIM/MQRGN	ERQAN/FLGKI	RPGNF/LQNRP
	Week 164	L	K	R	D	M	I	Q	V	A	N	VSQNY/PIVQN	KARVL/AEAMS	NANIM/MQRGN	ERQAN/FLGKI	RPGNF/LQNRP
Study cohort (n=10)	Baseline (n=10)	L	R	K	E	M	I	Q	V	G	N	VSQNY/PIVQN	KARVL/AEAMS	NTNIM/MQRSN	ERQAN/FLGKI	RPGNF/LQNRP
	Week 40 (n=8)	L	R	K	E	M	I	Q	V	G	N	VSQNY/PIVQN	KARVL/AEAMS	NTNIM/MQRSN	ERQAN/FLGKI	RPGNF/LQSRP

Table 20. Variation of Gag cleavage sites. Cleavage site sequences were compared to HXB2 amino acid positioning (Genbank accession number K035). In red are amino acids at positions that are known to confer reduced susceptibility to PI.

3.3.1.4.3 Variation of sites in Gag - p17, p24, p7 and p6

		Gag (non-cleavage site elements)																								
		Protease										p17 (MA)			p24 (CA)		p2		p7 (NC)				p6			
Amino acid position		10	14	20	35	46	54	58	82	94	98	76	79	81	200	219	370	371	389	390	401	409	468	474	487	497
HXB2		L	R	K	E	M	I	Q	V	G	N	R	Y	T	M	H	V	T	I	V	I	R	E	Q	T	P
134102	Baseline	L	R	R	E	M	I	Q	V	G	N	R	Y	T	M	H	V	-	T	V	I	R	E	Q	T	P
	Week 40	L	R	R	E	M	I	Q	V	G	N	R	Y	T	M	H	V	-	T	V	I	R	E	Q	T	P
	Week 96	L	R	R	E	M	I	Q	A	G	S	R	Y	T	M	H	V	-	T	V	L	R	E	Q	T	P
	Week 224	L	R	R	E	M	I	Q	A	G	N	R	Y	T	M	H	V	-	T	V	L	R	E	Q	T	P
	Week 298	F	R	R	E	I	V	E	A	G	N	R	Y	T	M	H	V	-	T	V	L	R	E	Q	T	S
141806	Baseline	L	R	K	E	M	I	Q	V	G	N	R	F	T	M	Q	V	-	I	V	I	R	E	Q	T	P
	Week 12	L	K	R	E	M	I	Q	V	G	N	R	F	T	M	Q	A	-	I	V	I	R	E	Q	I	P
	Week 40	L	K	R	D	M	I	Q	A	G	N	R	F	T	M	Q	A	-	I	V	I	R	E	Q	I	P
	Week 72	L	K	R	D	M	I	Q	A	G	N	R	F	T	M	Q	A	-	I	V	I	R	E	Q	I	P
	Week 96	L	K	R	D	M	I	Q	A	G	N	R	F	T	M	Q	A	-	I	V	I	R	E	Q	I	P
	Week 164	L	K	R	D	M	I	Q	V	A	N	R	F	T	M	Q	A	-	I	V	I	R	E	Q	I	P
Study cohort (n=10)	Baseline (n=10)	L	R	K	E	M	I	Q	V	G	N	R	F	T	M	Q	A	-	I	V	I	R	E	Q	I/T	P
	Week 40 (n=8)	L	R	K	E	M	I	Q	V	G	N	R	F	T	M	Q	A	-	I	V	I	R	E	Q	I	P

Table 21. Variation of Gag p17, p24, p7 and p6. Sites were compared to HXB2 amino acid positioning (Genbank accession number K035).

In red are amino acids at positions that are known to confer reduced susceptibility to PI. Key: ‘-’ is an amino acid deletion.

3.3.2 Discussion

In this chapter a comprehensive phenotypic assessment of viruses from two patients was undertaken. In published drug assays, comparable fitness in a single cycle replication assay have been reported between WT and Y181C mutated viruses, [252-254], in keeping with the data presented here. M184V is known to incur significant fitness cost to the virus compared to WT in the absence of 3TC [255, 256], which was also reflected in my RC data for single-mutated M184V pseudovirus from both children.

My PI-susceptibility results for V82A mutated pseudoviruses reflect what is already known about this mutation: V82A is known to confer between 2.8 and 16.2-fold change in IC50 compared to a reference strain without this mutation [205, 257, 258]. In combination with minor PI-resistance mutations in PR, the effect of V82A is enhanced to confer higher levels of resistance and cross-resistance to NFV and SQV [102]. Pseudovirus 4 from Study ID 141806 did not have any known major PI-resistance mutations in PR. M36I, R41K, L63T, I93L, K20R and E35D have all been described as common polymorphisms found in HIV-1 subtype C protease [188, 259, 260], so it was not a surprise to find them in the PR sequences of these children. Several studies have tried to assess the effect of these polymorphisms on the PI-susceptibility of HIV-1 subtype C viruses [188, 260-263]. My findings may implicate a R14K, K20R and E35D in PR to confer PI-resistance because these mutations emerged during PI-based therapy and they may also confer resistance in combination with one or more subtype C polymorphisms (M36I, R41K, L63P and I93L in PR). Future studies using site-directed mutagenesis to sequentially add or revert these

amino acids to the subtype-B amino acid should shed more light on the effect of these polymorphisms on the PI-susceptibility of subtype C viruses in these children. Based on my results, the recommended second-line regime (ABC, ddI and EFV) for participants in the CHER trial [78, 111] would be expected to successfully inhibit viral replication in the two children who developed multi-class drug resistance during ART. DRV would also be an option for children with a DRM profile like study ID 134102 at week 298 because this virus was resistant to all PIs available in paediatric formulations. The DRV:RTV ratio for paediatric formulations per age group also remains unclear but it is a priority formulation according to current WHO documentation [28].

The correlation coefficient as determined by Liu and Shafer [102], between M184V and reduced ddI susceptibility supports the theory that this mutation may confer low-level ddI resistance, however the impact of this mutation on ddI susceptibility remains conflicting. For example the loss of M184V has also been reported when 3TC is replaced by ddI in ART regimes [264] or this mutation has not been maintained in the presence of ddI [265]. My results reflect this conflict since I did not always observe reduced susceptibility to ddI with M184V mutated pseudovirus.

None of the amino acid changes observed in RT for this child have been described in the literature as associated with high-level ABC or ddI susceptibility. However, there were multiple background amino acids in RT that were different from the reference sequence provided by Stanford University HIV drug resistance database [102] that may be associated with high-level ABC and ddI resistance shown by Pseudovirus 3 from this child. These mutations in RT were V35T, E36D, T39R, S48T, T165I, K173A, D177E, T200A, Q207D, R211K, V245Q, E248N, D250N, A272P, K277R,

E291D, I293V, T296N and one or several of these mutations may be associated with high-level ABC and ddI resistance in combination with M184V. This requires further testing.

My results also confirm that persisting Y181C mutated virus at week 40 of PI-based ART (Pseudovirus 2 for Study ID 134102) continued to confer reduced susceptibility to NVP. I also expected this virus to have reduced susceptibility to EFV since Y181C is known to confer intermediate EFV resistance [102]. The variance of the mean EFV IC₅₀ from the two independent experiments for this drug with pseudovirus 2 was significantly different ($p=0.0075$), which explains this result.

P453L in the p1/p6 cleavage site of Gag is known to enhance PI resistance when in combination with I50V, I84V and L90M in PR [208, 266, 267] and the current literature has only described P453L selection by IDV [154]. Given the treatment history of this child is IDV-free and the lack of results from my literature review, this may be the first time that P453L in the p1/p6 cleavage site of Gag has been selected by LPV. On its own, P453L has not been shown to confer PI-resistance; instead it enhances resistance to all the PIs approved for clinical use if present in combination with I50V, I84V, N88D and L90M in PR (10 - 15 fold change in PI IC₅₀) [267], which can compromise the use of PIs in Study ID 134102. The effect of P453L in Gag on DRV susceptibility has not yet been determined.

The mechanism of action of the P453L mutation in the p1/p6 cleavage site of Gag in combination with the mentioned amino acid changes in protease is likely to be similar to that of the A431V mutation in the same cleavage site [268]: it can fill gaps between the substrate and the mutated protease enzyme, compensating for the loss of contact between the two when only the enzyme has mutated.

In the viral population of Study ID 141806, T487I in the p6 region of Gag emerged before V82A and simultaneously with K20R in PR at week 12 of ART. Therefore one could argue that this mutation may facilitate the emergence of the Major PI mutation, V82A, in PR. And/or T487I in the p6 of Gag could be a compensatory mutation to the minor PI mutation K20R in PR.

Usually mutations in Gag that are associated with PI-exposure and PI-resistance and the frequency of these amino acids have been determined for HIV-1 subtype B [154, 206] and determined in HIV-1 infected adults. Based on my review of the literature, this may be the first study that examines Gag variability in children vertically infected with HIV-1 subtype C. My PSSM analysis showed high conservation of amino acid sequence except in the p2/p7 cleavage site and a deletion at position 371 in p2 as most frequent in subtype C treatment naïve children from Sub-Saharan Africa, which was reflected in the sequences from the children in my study. I also found the deletion at position 371 to be the most frequent occurrence, which was also reflected in 100% of the sequences from my study. Future work on the phenotypic effect of these Gag polymorphisms and changes seen in the Gag in my patients would give insight into their effect on PI-susceptibility. This would be especially important for the children who were failing PI-based therapy in the absence of known PI- resistance conferring mutations in the viral protease, as well as to highlight the importance of HIV-1 subtype C Gag polymorphisms in children, akin to adult studies [154, 269, 270].

Introductory section 1.4.3.2 discussed what is known and thought about polymorphisms in the Gag cleavage sites and matrix that have been associated with PI-susceptibility may be more prevalent in non-subtype B viruses and my PSSM results suggest such a relationship for subtype C. It is also likely that the mutations

observed in Gag known to be associated with PI-exposure were linked on the same genome with known major mutations observed in protease, though I could not formally prove this. Therefore I also identify the need for my drug susceptibility and replication capacity assays to be repeated with pseudoviruses that contained patient gag-PR-RT from single genomes to determine the contribution of co-evolved sites in Gag to PI resistance and the relevance of subtype C polymorphisms in Gag.

3.3.3 Conclusions

The results of this chapter confirm and extend to children infected with subtype C HIV-1 what is known about the effect of known drug resistance mutations in PR and RT on susceptibilities and replication capacities. However the PI susceptibility of Pseudovirus 4 from Study ID 141806 adds to the discussion on the outcomes of polymorphisms in PR to PI-susceptibility. My results were in support of the possibility of R14K, K20R, E35D M36I, R41K, L63P and I93L in PR contributing to reduced PI-susceptibility and further work needs to be done to determine this. More work needs to be done on the effect of polymorphisms in Gag on PI-resistance, especially in treatment naïve children and the phenotypic contributions of genetic linkage of Gag mutations and polymorphisms to mutations in PR. Finally, I highlight the importance of the development of appropriate paediatric DRV co-formulations and dosages.

Chapter 4 Final Discussion, Project Limitations and Future

Work

Sub-Saharan Africa continues to have the highest burden of new paediatric HIV infections, but the availability and uptake of ART by HIV infected children has been increasing since 2005 [25]. An implication of the increased access ART is that more children eventually fail first-line treatment and the demand for second-line therapy drug regimens increases especially in more rural areas where patients are poorly monitored [91, 271]. In addition, access to second-line paediatric ART in resource-poor settings is very challenging and ARV options are limited [272].

Since 2008, WHO guidelines [81] stipulated that children who are confirmed as HIV positive start LPV/r-based therapy regardless of their immune or clinical status. This is because of previous exposure to NVP for PMTCT and the superiority of LPV/r-based first line therapy. One such superiority is LPV/r's relatively high genetic barrier to resistance compared to RTI-based treatment, that is more forgiving of suboptimal adherence to therapy (adherence <95%) [273]. However the effectiveness of this drug regime in children is confounded by metabolic complications [274], poor-palatability [274] and drug-drug interactions [275].

The South African CHER trial demonstrated a 76% (95% CI: 49%–89%) reduction in mortality associated with early ART compared to deferred ART in children <12 weeks of age [78, 93, 111]. The effects of early ART in older children are less clear [276-278] and evidence from randomized controlled trials regarding the optimal timing of ART initiation in children between 2 and 5 years of age [276, 277] is still

too deficient to influence treatment policy guidelines. While early ART may reduce mortality in infants, it could also increase the risk of toxicity and earlier development of drug resistance. South Africa also has a high burden of tuberculosis for which rifampicin is used for treatment in HIV infected children [279], but LPV/r - rifampicin interactions can lead to sub-therapeutic levels of LPV and virological failure due to the development of DRMs [275, 280-282].

HIV resistance testing is not generally available in resource-poor settings because of the lack of expertise required for data analysis and the expense of high-quality implementation and maintenance of this service [283]. Even though South Africa has a low prevalence rate of transmitted PI and NRTI resistance in children (<5%) [233, 284], transmitted NNRTI resistance is between 5% and 15% because of sdNVP for PMTCT [88, 284]. Therefore failure of first-line PI-based regimes because of the development of PI and/or NRTI resistance can severely limit ARV options for children in South Africa or similar resource-poor settings.

My aim was to describe the longitudinal development of drug resistance in the viral population of HIV-1 infected children who failed early ART in the CHER trial and explore the contribution of variants in the viral population to drug resistance and treatment failure. In particular, I theorized (1) the development of drug resistance mutations in PR and RT that were linked on the same genome, (2) the persistence of NNRTI resistance during PI-based ART and (3) drug resistant minority species contributing to future ART outcomes by eventually becoming the dominant species under the correct conditions.

I was able to use SGS to show multiple drug resistance mutations in PR and RT that were genetically linked. Viruses that harboured these genomes were shown to confer reduced susceptibility to 2 or 3 drug classes in my drug susceptibility assay. During early ART, two children developed the 3TC-selected resistance conferring mutation M184V, which conferred high-level resistance to 3TC (part of the first line ART regime) in two children as well as high-level cross-resistance to ABC, (a component of second line ART), in one of these children. V82A in PR was always selected in M184V mutated viruses, however my phylogenetic analyses did not determine the emergence of these mutations to be dependent on each other.

After treatment interruption and during a second round of recycled PI-based ART, dual-class drug resistant viruses persisted in the majority species of the viral population of one child and triple class drug resistant viruses were present in the minority species. This child received rifampicin for the treatment of tuberculosis and also had additional RTV to achieve a 1:1 ratio of LPV:RTV for ART in accordance with local HIV treatment guidelines [275]. This child developed multiple PI-resistance mutations that were genetically linked in PR as well as P453L in Gag at the population level. Therefore it is possible that additional RTV may not have been sufficient to prevent sub-therapeutic levels of LPV during rifampicin treatment, which caused the selection of multiple PI-resistance mutations in PR and Gag. Unfortunately I did not have sufficient sample volumes to determine the concentration of LPV circulating in this child at each sampling time point. I also did not have enough sample volumes to compare circulating LPV concentration in the other children from my study cohort.

Multiple studies have shown the correlation between virological failure during NNRTI-based ART and pre-existing minority NNRTI resistance [117, 123, 126-128, 285]. Several studies have also shown the correlation between virological failure during NNRTI-based ART and prior exposure to NVP for PMTCT in mothers [286] and children [89, 287]. These NNRTI resistant viruses are also likely to be archived in resting CD4+ T-cells [288, 289] that implies a permanent risk of the re-emergence of these viruses that can compromise future ART outcomes. The recommended second-line based ART regime in the CHER trial was EFV-based ART for children <3 years old or NVP-based ART for children >3 years old [78, 111]. I detected NNRTI resistance selected by NVP for PMTCT at baseline and to my knowledge I am the first to show that these viruses (K101E and Y181C mutated viruses) can persist in the viral population of children as minority species until week 40 of LPV/r-based ART. They are also likely to be present in the latent viral reservoir and have the potential to compromise EFV- or NVP-based second line ART. In one child, V108I was detected at week 96 of early ART and week 298 of recycled ART after treatment interruption. This child was >3 years old with virological failure on recycled LPV/r based ART and the V108I mutated virus detected at this time point conferred high-level resistance to NVP (35-fold decreased susceptibility to NVP compared the WT). Therefore this virus presents a risk to second-line ART failure for this child because of possible selection of circulating V108I-mutated viruses and V108I-mutated viruses assumed to be in the viral reservoir.

In my study, there were several instances of DRMs in the minority species of the viral populations of some children before first-line ART began and these viruses had the potential to be selected by components of early ART. These mutations were L74V,

T215C and K219N in RT and I50V in PR. Despite the first-line ART regime these mutations were not detected by SGS in the viral populations of these children during ART. The literature is divided on the contribution of pre-existing DRMs to the outcomes of first-line ART but my results are in-line with the body of evidence that does not correlate pre-existing DRMs in the minority species with virological failure during first-line ART [99, 290, 291].

To my knowledge, this project is also the first to identify multi-class drug resistance mutations in PR and RT that were linked on the same genome as well as characterise their development during early PI-based ART in children. All multi-class drug resistant viruses had M184V in RT and one or more PI-resistance mutations in PR.. The triple class drug resistant viruses detected in the minority species of the viral population of Study ID 134102 demonstrated significant levels of resistance to LPV, SQV, NFV, 3TC and NVP.

I also saw that early ART did not effectively eliminate the viral reservoir so that it continued to be a major contributor to the evolution of the viral populations during virological success and failure on ART; highlighting the importance of the viral reservoir.

This project was limited in several ways. First, There were a relatively low number of children who failed first-line PI-based early ART (27/377) from the CHER trial and the project used a convenience sample that of an even lower number of children (n = 10), which severely limited the power of the study. It was also unfortunate that my study did not include longitudinal samples from a control group, i.e. children who did not fail early ART from the CHER study. Insufficient volumes of plasma from each

child limited some prospects for my research: I was reliant on leftover plasma samples after viral load testing during the CHER trial, therefore I could not determine drug concentrations or obtain gag-PR-RT single genomes to determine genetic linkage of DRMs between these genes. I was also unable to reliably assess treatment adherence, I did not have access to plasma samples or treatment history of mothers. All of these limitations restricts the generalizability of this study so that I cannot infer outcomes for children receiving early ART or for children in other resource-limited settings. Notwithstanding these limitations, the study results suggests that prophylaxis and early ART may not just be strategies to improve child mortality and disease progression, but may also be strategies to restrict the viral reservoir as suggested recently by Kearney et al [222]. However it can also lead to multi-class drug resistance viruses that have the potential to severely limit ARV options, especially in resource-limited settings.

More research needs to be undertaken before the association between early initiation of ART and the risks of toxicity and development of drug resistance is more clearly understood in children. My findings support calls [292] to evaluate the effect of rifampicin concentrations on therapeutic levels of LPV/r with additional RTV and the development of drug resistance in children is needed. More prospective studies or a follow-up study from the CHER trial is needed to effectively evaluate the contribution of minority species detected during early ART to second-line treatment outcomes, including the contribution of genetic linkage of mutations in Gag, PR and RT and the evolution of DRMs in the latent viral reservoirs.

Chapter 5 Appendices

5.1 Appendix A

Patient ID	Time between DOB and baseline sample (weeks)	Time between start of early ART and baseline sample (weeks)
143646	5.1	1.0
141586	8.6	1.0
130166	6.4	0.4
138506	5.0	1.1
141806	5.6	1.0
131326	6.1	1.0
147636	5.4	1.0
134102	7.9	1.0
153716	7.6	1.0
146666	8.7	1.0

Table 22. Date of birth (DOB) and start of early ART relative to the time the baseline samples were taken for 10 children from the CHER study.

5.2 Appendix B

APOBEC3G/F cytidine deamination of PR-RT

I determined the contribution of cytidine deamination by APOBEC3G/F to the diversity of HIV-1 *pol* and the development of drug resistance in the viral populations of 9 children from my study cohort. I analysed the single genome sequences from all the sampling time points in each from a single MSA per child. Based on the ML trees of the viral population infecting each child, I used the most ancestral sequence obtained from each child as the reference sequence for these analyses and a built-in Fisher exact test detected any increases in APOBEC3G/F-mediated G-to-A mutations in the test sequences of the MSA compared to the reference sequence. Only one child had evidence of this type of mutation and it was a single sequence at week 164 of ART for Study ID 131326. This sequence had 10/197 (5%) sites that were APOBEC3G/F-mediated G-to-A mutations (Fisher exact test p-value = 0.00573719). I removed this sequence from the MSA for this child before continuing with further analyses.

	Study ID	Time point (Week)	MPD	Standard Deviation of MPD	Number of PR-RT single genome sequences obtained	Length of PR-RT (nucleotides)	Ntd changes in PR-RT based on MPD
Group 1	141586	0	0.0012	0.00101	30	1489	1
		24	-	-	-		
		40	0.0012	0.00099	39		1
	138506	0	0.0012	0.00123	57	1509	1
		24	-	-	-		
		40	0.0045	0.00177	40		6
	146666	0	0.001	0.00079	26	1477	1
		24	-	-	-		
		40	0.0018	0.00129	39		2
		72	-	-	-		
		96	0.003	0.00117	14		4
	147636	0	0.0014	0.00103	34	1507	1
		24	-	-	-		
		40	0.0008	0.00056	29		1
		72	-	-	-		
96		-	-	-			
164		0.005	0.00168	39	7		
Group 2	143646	0	0.0002	0.00034	49	1498	<1
		24	-	-	-		
		40	0.0001	0.00020	22		<1
	130166	0	0.0024	0.00332	49	1477	3
		24	-	-	-		
		40	0.0013	0.00066	25		1
	131326	0	0.0012	0.00117	21	1483	1
		24	-	-	-		

		40	0.0019	0.00083	38		2
		72	0.0039	0.00138	12		5
		96	-	-	-		
		164	0.0031	0.00198	39		4
Group 3	141806	0	0.0069	0.00477	33	1495	10
		12	0.0057	0.00515	37		8
		24	-	-	-		
		40	0.0046	0.00394	28		6
		72	0.002	0.00202	44		2
		96	0.0034	0.00215	30		5
		164	0.0075	0.00286	38		11
	224	0.0108	0.00011	22	16		
	134102	0	0.0016	0.00093	31	1498	2
		24	-	-	-		
		40	0.003	0.00148	38		4
		72	-	-	-		
		96	0.0038	0.00208	32		5
		164	-	-	-		
224		0.0028	0.00230	34	4		
298	0.0014	0.00119	39	2			

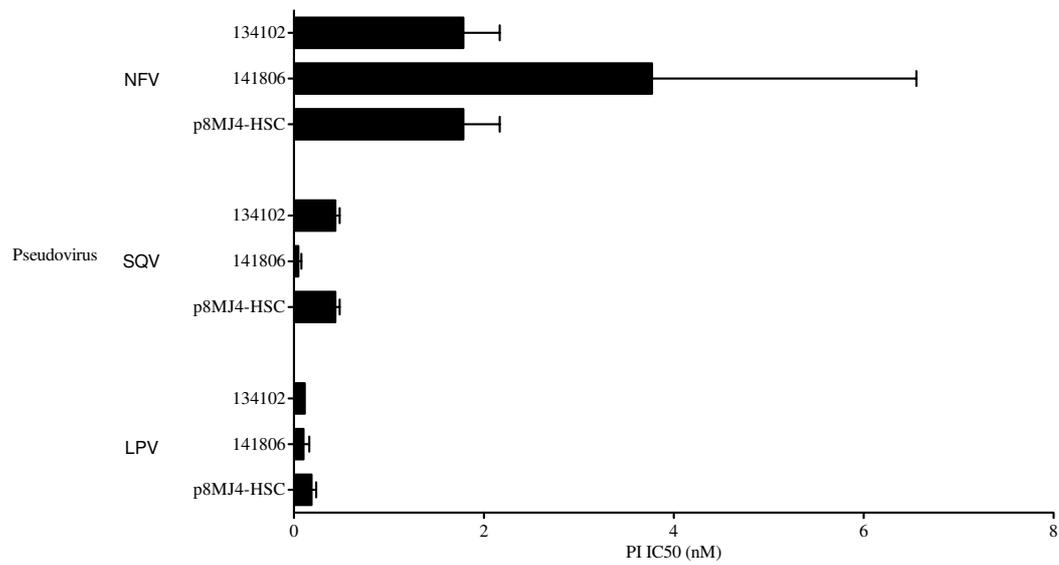
Table 23. Mean pairwise genetic distance (nucleotide substitutions per site in PR-RT) for 9 children. Also shown are the Standard Deviations of the MPDs, the number of single genome sequences obtained at each time point and the absolute number of nucleotide changes in PR-RT at each time point based on the MPD and the consensus nucleotide length of PR-RT obtained from each child.

5.3 Appendix C

Drug / drug class	Abbreviation
Nucleoside reverse transcriptase inhibitors	
Lamivudine	3TC
Zidovudine	AZT
Didanosine	ddI
Abacavir	ABC
Non-nucleoside reverse transcriptase inhibitors	
Efavirenz	EFV
Nevirapine	NVP
Protease inhibitors	
Ritonavir	RTV
Lopinavir	LPV
Lopinavir boosted with RTV	LPV/r
Darunavir	DRV
Tipranavir	TPV
Atazanavir	ATV

Table 24. List of antiretrovirals, their abbreviations and drug class.

(A)



(B)

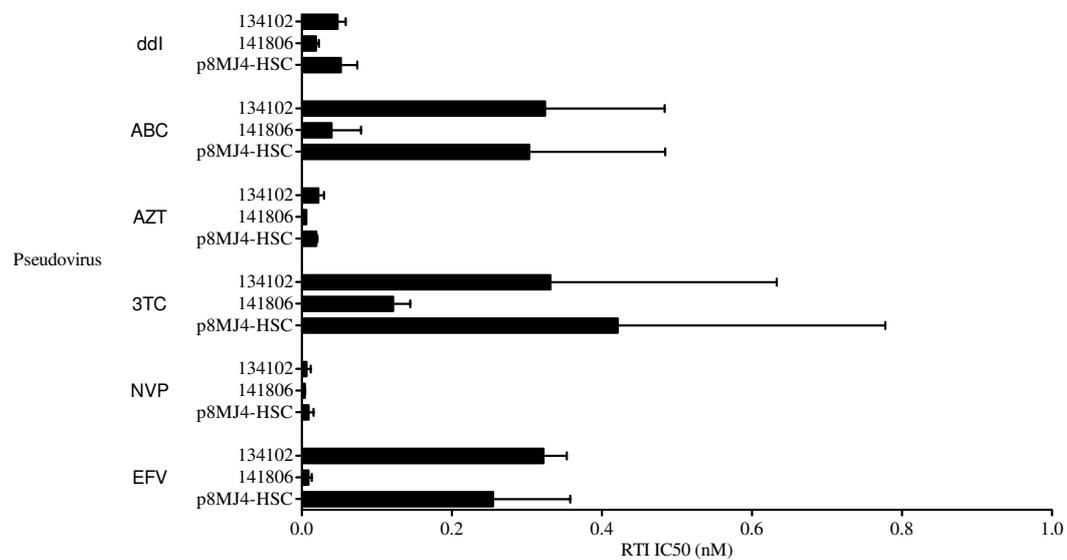
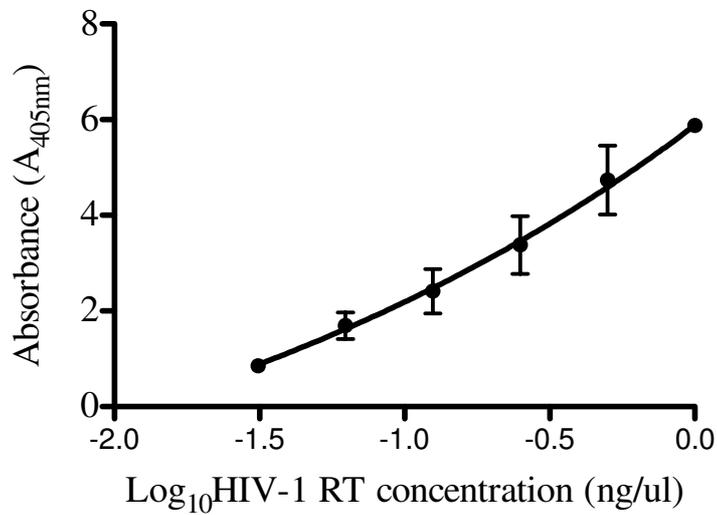


Figure 5.1 IC₅₀ of p8MJ4-HSC compared to the IC₅₀ of Pseudovirus 1 of Study ID 141806 and 134102. (A) PI (B) RTI. Student's t-determined that the mean IC₅₀ of 2 independent experiments for the patient pseudoviruses were not significantly different from p8MJ4-HSC for each drug ($p > 0.05$).

(A)



(B)

Study ID	Pseudovirus	HIV-1 RT concentration (ng/ul)
141806	WT	0.014
	184V	0.015
	184V+82A	0.015
	E4	0.014
134102	WT	0.014
	184V	0.014
	184V+82A	0.015
	184V+82A+54V+58E+46I+10F	0.015
	108I+184V+82A+54V+58E+46I+10F	0.016
	p8MJ4-HSC	0.014

Figure 5.2 Standardization of pseudoviral titres. (A) HIV-1 RT ELISA calibration curve. (B) HIV-1 RT concentrations from pseudovirus produced from recombinant gag-pol/patient PR-RT expression vectors. p8MJ4-HSC was the original expression vector.

References

1. Prevention CfDCa. Current Trends Update on Acquired Immune Deficiency Syndrome (AIDS) -- United States. *Morbidity and Mortality Weekly* 1982,**31**:507-508.
2. F B-S, JC C, F R, MT N, S C, J G, *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983,**220**:868-871.
3. JA L, AD H, SM K, JA L, JM S, LS O. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* 1984,**225**:840-842.
4. Coffin J, Haase A, Levy J, Montagnier L, Oroszlan S, Teich N, *et al.* What to call the AIDS virus? *Nature* 1986,**321**:10.
5. 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**:1789-1796.
6. 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**:436-441.
7. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *Aids* 2006,**20**:W13-W23.
8. Vallari A, Bodelle P, Ngansop C, Makamche F, Ndembi N, Mbanya D, *et al.* Four new HIV-1 group N isolates from Cameroon: Prevalence continues to be low. *AIDS Res Hum Retroviruses* 2010,**26**:109-115.

9. Peeters M, Gueye A, Mboup S, Bibollet-Ruche F, Ekaza E, Mulanga C, *et al.* Geographical distribution of HIV-1 group O viruses in Africa. *Aids* 1997,**11**:493-498.
10. Plantier J-C, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, *et al.* A new human immunodeficiency virus derived from gorillas. *Nature medicine* 2009,**15**:871-872.
11. Hemelaar J Fau - Gouws E, Gouws E Fau - Ghys PD, Ghys Pd Fau - Osmanov S, Osmanov S. Global trends in molecular epidemiology of HIV-1 during 2000-2007.
12. Vidal N, Mulanga C, Bazepeo SE, Mwamba JK, Tshimpaka J-W, Kashi M, *et al.* Distribution of HIV-1 variants in the Democratic Republic of Congo suggests increase of subtype C in Kinshasa between 1997 and 2002. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2005,**40**:456-462.
13. Soares EA, Martínez AM, Souza TM, Santos AF, Da Hora V, Silveira J, *et al.* HIV-1 subtype C dissemination in southern Brazil. *Aids* 2005,**19**:S81-S86.
14. Luo C-C, Tian C, Hu D, Kai M, Dondero T, Zheng X. HIV-1 subtype C in China. *The Lancet* 1995,**345**:1051-1052.
15. Piyasirisilp S, McCutchan FE, Carr JK, Sanders-Buell E, Liu W, Chen J, *et al.* A recent outbreak of human immunodeficiency virus type 1 infection in southern China was initiated by two highly homogeneous, geographically separated strains, circulating recombinant form AE and a novel BC recombinant. *J Virol* 2000,**74**:11286-11295.
16. Hahn BH, Shaw GM, De KM, Sharp PM. AIDS as a zoonosis: scientific and public health implications. *Science* 2000,**287**:607-614.

17. Kanki P, De Cock K. Epidemiology and natural history of HIV-2. *AIDS-LONDON-CURRENT SCIENCE THEN RAPID SCIENCE PUBLISHERS THEN LIPPINCOTT RAVEN*- 1994,**8**:S85-S85.
18. Control CfD. Factsheet HIV Type 2. 2010.
19. Gao F, Yue L, Robertson DL, Hill SC, Hui H, Biggar RJ, *et al.* Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology. *J Virol* 1994,**68**:7433-7447.
20. Chen Z, Luckay A, Sodora DL, Telfer P, Reed P, Gettie A, *et al.* Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. *J Virol* 1997,**71**:3953-3960.
21. Aguchi JY, Devare SG, Brennan CA. Sequence Note: Identification of a New HIV-2 Subtype Based on Phylogenetic Analysis of Full-Length Genomic Sequence. *AIDS Res Hum Retroviruses* 2000,**16**:925-930.
22. Damond F, Worobey M, Campa P, Farfara I, Colin G, Matheron S, *et al.* Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification. *AIDS Res Hum Retroviruses* 2004,**20**:666-672.
23. Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme A-M. Tracing the origin and history of the HIV-2 epidemic. *Proceedings of the National Academy of Sciences* 2003,**100**:6588-6592.
24. Hemelaar J. The origin and diversity of the HIV-1 pandemic. *Trends in molecular medicine* 2012,**18**:182-192.

25. UNAIDS. UNAIDS report on the global AIDS epidemic. *UNAIDS Global Report* 2012.
26. Shannon KM, Ammann AJ. Acquired immune deficiency syndrome in childhood. *The Journal of pediatrics* 1985,**106**:332-342.
27. Organization WH. HIV/AIDS Data Products: Size of the Epidemic. *Global Health Observatory* 2011.
28. Organization WH. Antiretroviral therapy of HIV infection in infants and children: towards universal access. 2010,**Annex E**.
29. UNAIDS. The Gap Report. 2013.
30. Paintsil E, Andiman WA. Care and management of the infant of the HIV-1-infected mother. In: *Seminars in perinatology*: Elsevier; 2007. pp. 112-123.
31. De Cock K, Fowler MG, Mercier E, de Vincenzi I, Saba J, Hoff E, *et al*. Prevention of mother-to-child HIV transmission in resource-poor countries: translating research into policy and practice. *JAMA*. 2000,**283**:1175-1182.
32. Lehman DA, Farquhar C. Biological mechanisms of vertical human immunodeficiency virus (HIV - 1) transmission. *Reviews in medical virology* 2007,**17**:381-403.
33. Dunn D, Newell M, Mayaux M, Kind C, Goedert J, Andiman W, *et al*. Mode of delivery and vertical transmission of HIV-1: a review of prospective studies. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 1994,**7**:1064-1066.
34. Dunn DT, Newell M-L, Ades A, Peckham CS. Risk of human immunodeficiency virus type 1 transmission through breastfeeding. *The Lancet* 1992,**340**:585-588.

35. UNAIDS. UNGASS Declaration: Resolution adopted by the General Assembly; Declaration of Commitment on HIV/AIDS. 2001.
36. Taha TE, Kumwenda NI, Gibbons A, Broadhead RL, Fiscus S, Lema V, *et al.* Short postexposure prophylaxis in newborn babies to reduce mother-to-child transmission of HIV-1: NVAZ randomised clinical trial. *The Lancet* 2003,**362**:1171-1177.
37. Taha TE, Kumwenda NI, Hoover DR, Fiscus SA, Kafulafula G, Nkhoma C, *et al.* Nevirapine and zidovudine at birth to reduce perinatal transmission of HIV in an African setting: a randomized controlled trial. *Jama* 2004,**292**:202-209.
38. Guay LA, Musoke P, Fleming T, Bagenda D, Allen M, Nakabiito C, *et al.* Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial. *The Lancet* 1999,**354**:795-802.
39. Organization WH. Prevention of mother-to-child transmission of HIV: Selection and use of Nevirapine - Technical Notes. *Geneva* 2001.
40. Jackson JB, Musoke P, Fleming T, Guay LA, Bagenda D, Allen M, *et al.* Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: 18-month follow-up of the HIVNET 012 randomised trial. *The Lancet* 2003,**362**:859-868.
41. Bedri A, Gudetta B, Isehak A, Kumbi S, Lulseged S, Mengistu Y, *et al.* Extended-dose nevirapine to 6 weeks of age for infants to prevent HIV transmission via breastfeeding in Ethiopia, India, and Uganda: an analysis of three randomised controlled trials. *Lancet* 2008,**372**:300-313.

42. Kumwenda NI, Hoover DR, Mofenson LM, Thigpen MC, Kafulafula G, Li Q, *et al.* Extended antiretroviral prophylaxis to reduce breast-milk HIV-1 transmission. *New England Journal of Medicine* 2008,**359**:119-129.
43. Chasela CS, Hudgens MG, Jamieson DJ, Kayira D, Hosseinipour MC, Kourtis AP, *et al.* Maternal or infant antiretroviral drugs to reduce HIV-1 transmission. *New England Journal of Medicine* 2010,**362**:2271-2281.
44. (WHO) WHO. 2010 Antiretroviral drugs for treating pregnant women and preventing HIV infection in infants in resource-limited settings: Towards universal access. Recommendations for a public health approach. *WHO HIV PMTCT treatment guidelines* 2010.
45. Organization WH. Use of antiretroviral drugs for treating pregnant women and preventing HIV infection in infants. *World Health Organization Programmatic Update* 2013,**WHO/HIV/2012.6**.
46. WHO, UNICEF, UNAIDS. Progress report 2011: Global HIV/AIDS response. 2011.
47. HIV/AIDS WHO. Treatment of children living with HIV: Latest updates. 2011.
48. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, *et al.* Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine* 2011,**365**:493-505.
49. Shapiro RL, Thior I, Gilbert PB, Lockman S, Wester C, Smeaton LM, *et al.* Maternal single-dose nevirapine versus placebo as part of an antiretroviral strategy to prevent mother-to-child HIV transmission in Botswana. *Aids* 2006,**20**:1281-1288.

50. Thior I, Lockman S, Smeaton LM, Shapiro RL, Wester C, Heymann SJ, *et al.* Breastfeeding plus infant zidovudine prophylaxis for 6 months vs formula feeding plus infant zidovudine for 1 month to reduce mother-to-child HIV transmission in Botswana: a randomized trial: the Mashi Study. *Jama* 2006,**296**:794-805.
51. Dabis F, Bequet L, Ekouevi DK, Viho I, Rouet F, Horo A, *et al.* Field efficacy of zidovudine, lamivudine and single-dose nevirapine to prevent peripartum HIV transmission. *AIDS (London, England)* 2005,**19**:309.
52. Connor EM, Sperling RS, Gelber R, Kiselev P, Scott G, O'Sullivan MJ, *et al.* Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. *New England Journal of Medicine* 1994,**331**:1173-1180.
53. Shaffer N, Chuachoowong R, Mock PA, Bhadrakom C, Siriwasin W, Young NL, *et al.* Short-course zidovudine for perinatal HIV-1 transmission in Bangkok, Thailand: a randomised controlled trial. *The Lancet* 1999,**353**:773-780.
54. Baggaley RF, White RG, Boily M-C. Systematic review of orogenital HIV-1 transmission probabilities. *Int J Epidemiol* 2008,**37**:1255-1265.
55. Hladik F, Hope TJ. HIV infection of the genital mucosa in women. *Current HIV/AIDS Reports* 2009,**6**:20-28.
56. Pope M, Haase AT. Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. *Nature medicine* 2003,**9**:847-852.
57. Winkelstein W, Lyman DM, Padian N, Grant R, Samuel M, Wiley JA, *et al.* Sexual practices and risk of infection by the human immunodeficiency virus: the San Francisco Men's Health Study. *Jama* 1987,**257**:321-325.

58. Health JHBSOP. AIDS Linked to the Intravenous Experience studies - Major contributions. In; 1988 - 2013.
59. Mofenson LM. Advances in the prevention of vertical transmission of human immunodeficiency virus. In: *Seminars in pediatric infectious diseases*: Elsevier; 2003. pp. 295-308.
60. Coffin J, Swanstrom R. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harbor perspectives in medicine* 2013,**3**:a012526.
61. Tovo P, Gabiano C, Palomba E, De Martino M, Galli L, Cappello N, *et al.* Prognostic factors and survival in children with perinatal HIV-1 infection. *The Lancet* 1992,**339**:1249-1253.
62. Blanche S, Newell M-L, Mayaux M-J, Dunn DT, Teglas JP, Rouzioux C, *et al.* Morbidity and mortality in European children vertically infected by HIV-1: the French Pediatric HIV Infection Study Group and European Collaborative Study. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 1997,**14**:442-450.
63. Blanche S, Tardieu M, Duliege A-M, Rouzioux C, Le Deist F, Fukunaga K, *et al.* Longitudinal study of 94 symptomatic infants with perinatally acquired human immunodeficiency virus infection: evidence for a bimodal expression of clinical and biological symptoms. *American Journal of Diseases of Children* 1990,**144**:1210-1215.
64. Scott GB, Hutto C, Makuch RW, Mastrucci MT, O'Connor T, Mitchell CD, *et al.* Survival in children with perinatally acquired human immunodeficiency virus type 1 infection. *New England Journal of Medicine* 1989,**321**:1791-1796.

65. Auger I, Thomas P, De Gruttola V, Morse D, Moore D, Williams R, *et al.* Incubation periods for paediatric AIDS patients. 1988.
66. Tobin NH, Aldrovandi GM. Immunology of pediatric HIV infection. *Immunological reviews* 2013,**254**:143-169.
67. Mofenson LM, Korelitz J, Meyer WA, Bethel J, Rich K, Pahwa S, *et al.* The relationship between serum human immunodeficiency virus type 1 (HIV-1) RNA level, CD4 lymphocyte percent, and long-term mortality risk in HIV-1—infected children. *Journal of Infectious Diseases* 1997,**175**:1029-1038.
68. Mayaux M-J, Burgard M, Teglas J-P, Cottalorda J, Krivine A, Simon F, *et al.* Neonatal characteristics in rapidly progressive perinatally acquired HIV-1 disease. *Jama* 1996,**275**:606-610.
69. Diaz C, Hanson C, Cooper ER, Read JS, Watson J, Mendez HA, *et al.* Disease progression in a cohort of infants with vertically acquired HIV infection observed from birth: the Women and Infants Transmission Study (WITS). *JAIDS Journal of Acquired Immune Deficiency Syndromes* 1998,**18**:221-228.
70. Dickover RE, Dillon M, Gillette SG, Deveikis A, Keller M, Plaeger-Marshall S, *et al.* Rapid increases in load of human immunodeficiency virus correlate with early disease progression and loss of CD4 cells in vertically infected infants. *Journal of Infectious Diseases* 1994,**170**:1279-1284.
71. Dickover RE, Dillon M, Leung K-M, Krogstad P, Plaeger S, Kwok S, *et al.* Early prognostic indicators in primary perinatal human immunodeficiency virus type 1 infection: importance of viral RNA and the timing of transmission on long-term outcome. *Journal of Infectious Diseases* 1998,**178**:375-387.
72. Kuhn L, Steketee RW, Weedon J, Abrams EJ, Lambert G, Bamji M, *et al.* Distinct risk factors for intrauterine and intrapartum human immunodeficiency

- virus transmission and consequences for disease progression in infected children. *Journal of Infectious Diseases* 1999,**179**:52-58.
73. Rich KC, Fowler MG, Mofenson LM, Abboud R, Pitt J, Diaz C, *et al.* Maternal and infant factors predicting disease progression in human immunodeficiency virus type 1-infected infants. *Pediatrics* 2000,**105**:e8-e8.
74. Chen Y, Winchester R, Korber B, Gagliano J, Bryson Y, Hutto C, *et al.* Influence of HLA alleles on the rate of progression of vertically transmitted HIV infection in children: association of several HLA-DR13 alleles with long-term survivorship and the potential association of HLA-A2301 with rapid progression to AIDS. *Human immunology* 1997,**55**:154-162.
75. Children IRfHI. Features of children perinatally infected with HIV-1 surviving longer than 5 years. *The Lancet* 1994,**343**:191-195.
76. Dunn D. HIV Paediatric Prognostic Markers Collaborative Study: Short-term risk of disease progression in HIV-1-infected children receiving no antiretroviral therapy or zidovudine monotherapy: a meta-analysis. *Lancet* 2003,**362**:1605-1611.
77. Luzuriaga K, McManus M, Mofenson L, Britto P, Graham B, Sullivan JL. A trial of three antiretroviral regimens in HIV-1-infected children. *New England Journal of Medicine* 2004,**350**:2471-2480.
78. Violari A, Cotton MF, Gibb DM, Babiker AG, Steyn J, Madhi SA, *et al.* Early Antiretroviral Therapy and Mortality among HIV-Infected Infants. *New England Journal of Medicine* 2008,**359**:2233-2244.
79. Organization WH. Antiretroviral therapy of HIV infection in infants and children in resource-limited settings: towards universal access: Recommendations for a public health approach. 2006.

80. Organization WH. Antiretroviral drugs for treating pregnant women and preventing HIV infection in infants Towards universal access: Recommendations for a public health approach (2006 revision) 2006.
81. Group WHOTR. Paediatric HIV/ART Care Guideline Group Meeting. *WHO Antiretroviral Therapy for Infants and Children* 2008.
82. Gupta RK, Gibb DM, Pillay D. Management of paediatric HIV-1 resistance. *CUrrent Opinions in Infectious Disease* 2009,**22**:256-263.
83. Sigaloff KC, Calis JC, Geelen SP, van Vugt M, de Wit TFR. HIV-1-resistance-associated mutations after failure of first-line antiretroviral treatment among children in resource-poor regions: a systematic review. *The Lancet infectious diseases* 2011,**11**:769-779.
84. Organization WH. Antiretroviral therapy of HIV infection in infants and children Towards universal access: Recommendations for a public health approach. 2006.
85. Turkova A, Webb RH, Lyall H. When to start, what to start and other treatment controversies in pediatric HIV infection. *Pediatric Drugs* 2012,**14**:361-376.
86. Organization WH. Antiretroviral Therapy for HIV Infection in Infants and Children Towards Universal Access: Recommendations for a public health approach. 2010 revision.
87. Organisation WH. Antiretrovirals in Pediatric Patients: Focus on Young Children: Appropriate Dosing, Adverse Effects, Efficacy. 2013.
88. Arrive E, Newell ML, Ekouevi DK, Chaix ML, Thiebaut R, Masquelier B, *et al.* Prevalence of resistance to nevirapine in mothers and children after single-

- dose exposure to prevent vertical transmission of HIV-1: a meta-analysis. *Int J Epidemiol* 2007,**36**:1009-1021.
89. Palumbo P, Lindsey JC, Hughes MD, Cotton MF, Bobat R, Meyers T, *et al.* Antiretroviral treatment for children with peripartum nevirapine exposure. *New England Journal of Medicine* 2010,**363**:1510-1520.
90. Kuhn L, Coovadia A, Strehlau R, Martens L, Hu C-C, Meyers T, *et al.* Switching children previously exposed to nevirapine to nevirapine-based treatment after initial suppression with a protease-inhibitor-based regimen: long-term follow-up of a randomised, open-label trial. *The Lancet infectious diseases* 2012,**12**:521-530.
91. Davies M-A, Moultrie H, Eley B, Rabie H, Van Cutsem G, Giddy J, *et al.* Virologic failure and second-line antiretroviral therapy in children in South Africa-The IeDEA Southern Africa Collaboration. *Journal of acquired immune deficiency syndromes (1999)* 2011,**56**:270.
92. Palumbo P, Violari A, Lindsey J, Hughes M, Jean-Philippe P, Mofenson L. NVP-vs LPV/r-based ART among HIV+ infants in resource-limited settings: The IMPAACT P1060 Trial. In: *18th Conference on Retroviruses and Opportunistic Infections (CROI)*; 2011.
93. Penazzato M, Prendergast A, Tierney J, Cotton M, Gibb D. Effectiveness of antiretroviral therapy in HIV-infected children under 2 years of age. *Cochrane Database Syst Rev* 2012,**7**:CD004772.
94. Violari A, Lindsey JC, Hughes MD, Mujuru HA, Barlow-Mosha L, Kamthunzi P, *et al.* Nevirapine versus ritonavir-boosted lopinavir for HIV-infected children. *New England Journal of Medicine* 2012,**366**:2380-2389.

95. Lindsey JC, Hughes MD, Violari A, Eshleman SH, Abrams EJ, Bwakura-Dangarembizi M, *et al.* Predictors of Virologic and Clinical Response to Nevirapine versus Lopinavir/Ritonavir-Based Antiretroviral Therapy in Young Children with and without Prior Nevirapine Exposure for the Prevention of Mother-To-Child HIV Transmission. *The Pediatric Infectious Disease Journal* 2014.
96. de Mulder M, Yebra G, Navas A, de José MI, Gurbindo MD, González-Tomé MI, *et al.* High drug resistance prevalence among vertically HIV-infected patients transferred from pediatric care to adult units in Spain. *PLoS ONE* 2012,**7**:e52155.
97. Wong FL, Hsu AJ, Pham PA, Siberry GK, Hutton N, Agwu AL. Antiretroviral treatment strategies in highly treatment experienced perinatally HIV-infected youth. *The Pediatric Infectious Disease Journal* 2012,**31**:1279.
98. Europe PLTOIptftCoOHER. Risk of triple-class virological failure in children with HIV: a retrospective cohort study. *Lancet* 2011,**377**:1580.
99. Hamers RL, Sigaloff KC, Wensing AM, Wallis CL, Kityo C, Siwale M, *et al.* Patterns of HIV-1 drug resistance after first-line antiretroviral therapy (ART) failure in 6 sub-Saharan African countries: implications for second-line ART strategies. *Clinical infectious diseases* 2012:cis254.
100. Odaibo GN, Okonkwo P, Adewole IF, Olaleye DO. Pattern of HIV-1 Drug Resistance among Adults on ART in Nigeria. *World Journal of AIDS* 2013,**3**:327.
101. Murphy RA, Sunpath H, Lu Z, Chelin N, Losina E, Gordon M, *et al.* Outcomes after virologic failure of first-line ART in South Africa. *AIDS (London, England)* 2010,**24**:1007.

102. Liu TF, Shafer RW. Web resources for HIV type 1 genotypic-resistance test interpretation. *Clinical infectious diseases* 2006,**42**:1608.
103. Brenner BG. Resistance and viral subtypes: how important are the differences and why do they occur? *Current Opinion in HIV and AIDS* 2007,**2**:94-102.
104. Martínez-Cajas JL, Pant-Pai N, Klein MB, Wainberg MA. Role of genetic diversity amongst HIV-1 non-B subtypes in drug resistance: a systematic review of virologic and biochemical evidence. *AIDS rev* 2008,**10**:212-223.
105. Grossman Z, Vardinon N, Chemtob D, Alkan ML, Bentwich Z, Burke M, *et al.* Genotypic variation of HIV-1 reverse transcriptase and protease: comparative analysis of clade C and clade B. *Aids* 2001,**15**:1453-1460.
106. Brenner B, Turner D, Oliveira M, Moisi D, Detorio M, Carobene M, *et al.* A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to non-nucleoside reverse transcriptase inhibitors. *Aids* 2003,**17**:F1-F5.
107. Loemba H, Brenner B, Parniak MA, Ma'ayan S, Spira B, Moisi D, *et al.* Genetic divergence of human immunodeficiency virus type 1 Ethiopian clade C reverse transcriptase (RT) and rapid development of resistance against nonnucleoside inhibitors of RT. *Antimicrobial agents and chemotherapy* 2002,**46**:2087-2094.
108. Soares RO, Batista PR, Costa MG, Dardenne LE, Pascutti PG, Soares MA. Understanding the HIV-1 protease nelfinavir resistance mutation D30N in subtypes B and C through molecular dynamics simulations. *Journal of Molecular Graphics and Modelling* 2010,**29**:137-147.
109. Gonzalez LM, Brindeiro RM, Aguiar RS, Pereira HS, Abreu CM, Soares MA, *et al.* Impact of nelfinavir resistance mutations on in vitro phenotype, fitness,

- and replication capacity of human immunodeficiency virus type 1 with subtype B and C proteases. *Antimicrobial agents and chemotherapy* 2004,**48**:3552-3555.
110. Violari A, Cotton M, Gibb D, Babiker A, Steyn J, Jean-Philippe P, *et al.* ART initiated before 12 weeks reduces early mortality in young HIV-infected infants: evidence from the Children with HIV Early Antiretroviral Therapy (CHER) Study. In: *4th International AIDS Society Conference on HIV Pathogenesis, Treatment and Prevention*. Australia; 2007.
 111. Cotton MF, Violari A, Otwoombe K, Panchia R, Dobbels E, Rabie H, *et al.* Early time-limited antiretroviral therapy versus deferred therapy in South African infants infected with HIV: results from the children with HIV early antiretroviral (CHER) randomised trial. *The Lancet* 2013,**382**:1555-1563.
 112. Services USDoHaH. The Division of AIDS Table for Grading the Severity of Adult and Pediatric Adverse Events Version 1.0. *National Institute of Allergy and Infectious Diseases* 2004.
 113. Halvas EK, Aldrovandi GM, Balfe P, Beck IA, Boltz VF, Coffin JM, *et al.* Blinded, multicenter comparison of methods to detect a drug-resistant mutant of human immunodeficiency virus type 1 at low frequency. *Journal of clinical microbiology* 2006,**44**:2612-2614.
 114. Jordan MR, Kearney M, Palmer S, Shao W, Maldarelli F, Coakley EP, *et al.* Comparison of standard PCR/cloning to single genome sequencing for analysis of HIV-1 populations. *Journal of virological methods* 2010,**168**:114-120.
 115. 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**:406-413.
116. Codoner FM, Pou C, Thielen A, Garcia F, Delgado R, Dalmau D, *et al.* Added value of deep sequencing relative to population sequencing in heavily pre-treated HIV-1-infected subjects. *PLoS ONE* 2011,**6**:e19461.
117. Coovadia A, Hunt G, Abrams EJ, Sherman G, Meyers T, Barry G, *et al.* Persistent minority K103N mutations among women exposed to single-dose nevirapine and virologic response to nonnucleoside reverse-transcriptase inhibitor-based therapy. *Clinical infectious diseases* 2009,**48**:462-472.
118. Eshleman S, Guay L, Mwatha A, Cunningham S, Brown E, Musoke P, *et al.* Comparison of nevirapine (NVP) resistance in Ugandan women 7 days vs. 6-8 weeks after single-dose NVP prophylaxis: HIVNET 012. *AIDS Res Hum Retroviruses* 2004,**20**:595-599.
119. Eshleman SH, Mracna M, Guay LA, Deseyve M, Cunningham S, Mirochnick M, *et al.* Selection and fading of resistance mutations in women and infants receiving nevirapine to prevent HIV-1 vertical transmission (HIVNET 012). *Aids* 2001,**15**:1951.
120. Flys T, Nissley DV, Claasen CW, Jones D, Shi C, Guay LA, *et al.* Sensitive drug-resistance assays reveal long-term persistence of HIV-1 variants with the K103N nevirapine (NVP) resistance mutation in some women and infants after the administration of single-dose NVP: HIVNET 012. *Journal of Infectious Diseases* 2005,**192**:24-29.
121. Ian Towler W, Church JD, Eshleman JR, Fowler MG, Guay LA, Jackson JB, *et al.* Analysis of Nevirapine Resistance Mutations in Cloned HIV Type 1 Variants from HIV-Infected Ugandan Infants Using a Single-Step

- Amplification-Sequencing Method (AmpliSeq). *AIDS Res Hum Retroviruses* 2008,**24**:1209-1213.
122. Metzner KJ, Rauch P, Walter H, Boesecke C, Zöllner B, Jessen H, *et al.* Detection of minor populations of drug-resistant HIV-1 in acute seroconverters. *Aids* 2005,**19**:1819-1825.
123. Metzner KJ, Giulieri SG, Knoepfel SA, Rauch P, Burgisser P, Yerly S, *et al.* Minority quasispecies of drug-resistant HIV-1 that lead to early therapy failure in treatment-naive and-adherent patients. *Clinical infectious diseases* 2009,**48**:239-247.
124. Metzner KJ, Bonhoeffer S, Fischer M, Karanickolas R, Allers K, Joos B, *et al.* Emergence of minor populations of human immunodeficiency virus type 1 carrying the M184V and L90M mutations in subjects undergoing structured treatment interruptions. *Journal of Infectious Diseases* 2003,**188**:1433-1443.
125. 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. *Journal of Infectious Diseases* 2004,**189**:1091-1096.
126. 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**:37-42.
127. Li Jz PRRHJ, *et al.* Low-frequency hiv-1 drug resistance mutations and risk of nrti-based antiretroviral treatment failure: A systematic review and pooled analysis. *JAMA: The Journal of the American Medical Association* 2011,**305**:1327-1335.

128. Paredes R, Lalama CM, Ribaud HJ, Schackman BR, Shikuma C, Giguél F, *et al.* Pre-existing minority drug-resistant HIV-1 variants, adherence, and risk of antiretroviral treatment failure. *Journal of Infectious Diseases* 2010,**201**:662-671.
129. Armenia D, Vandenbroucke I, Fabeni L, Van Marck H, Cento V, D'Arrigo R, *et al.* Study of genotypic and phenotypic HIV-1 dynamics of integrase mutations during raltegravir treatment: a refined analysis by ultra-deep 454 pyrosequencing. *Journal of Infectious Diseases* 2012,**205**:557-567.
130. Winters MA, Lloyd RM, Jr., Shafer RW, Kozal MJ, Miller MD, Holodniy M. Development of Elvitegravir Resistance and Linkage of Integrase Inhibitor Mutations with Protease and Reverse Transcriptase Resistance Mutations. *PLoS ONE* 2012,**7**:e40514.
131. Liu J, Miller MD, Danovich RM, Vandergrift N, Cai F, Hicks CB, *et al.* Analysis of low-frequency mutations associated with drug resistance to raltegravir before antiretroviral treatment. *Antimicrobial agents and chemotherapy* 2011,**55**:1114-1119.
132. Halvas EK, Wiegand A, Boltz VF, Kearney M, Nissley D, Wantman M, *et al.* Low frequency NNRTI-resistant variants contribute to failure of efavirenz-containing regimens in treatment-experienced patients. *J Infect Dis* 2010,**201**:672.
133. Cozzi-Lepri A, Noguera-Julian M, Di Giallonardo F, Schuurman R, Däumer M, Aitken S, *et al.* Low-frequency drug-resistant HIV-1 and risk of virological failure to first-line NNRTI-based ART: a multicohort European case-control study using centralized ultrasensitive 454 pyrosequencing. *Journal of Antimicrobial Chemotherapy* 2014:dku426.

134. 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. *J Clin Microbiol* 2005,**43**:406-413.
135. Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome research* 2007,**17**:1195-1201.
136. Van Laethem K, Van Vaerenbergh K, Schmit J-C, Sprecher S, Hermans P, De Vroey V, *et al.* Phenotypic assays and sequencing are less sensitive than point mutation assays for detection of resistance in mixed HIV-1 genotypic populations. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 1999,**22**:107-118.
137. Gibson RM, Schmotzer CL, Quiñones-Mateu ME. Next-Generation Sequencing to Help Monitor Patients Infected with HIV: Ready for Clinical Use? *Current infectious disease reports* 2014,**16**:1-9.
138. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences* 2008,**105**:7552-7557.
139. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, *et al.* Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J Virol* 2008,**82**:3952-3970.
140. Coffin JM, Hughes SH, E VH. Retroviruses. In. Edited by Press CSHL. New York: Cold Spring Harbour; 1997.

141. Li H, Dou J, Ding L, Spearman P. Myristoylation is required for human immunodeficiency virus type 1 Gag-Gag multimerization in mammalian cells. *J Virol* 2007,**81**:12899-12910.
142. Wiegers K, Rutter G, Kottler H, Tessmer U, Hohenberg H, Kräusslich HG. Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J Virol* 1998,**72**:2846-2854.
143. Moulard M, Decroly E. Maturation of HIV envelope glycoprotein precursors by cellular endoproteases. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* 2000,**1469**:121-132.
144. Verhoef K, Sanders RW, Fontaine V, Kitajima S, Berkhout B. Evolution of the human immunodeficiency virus type 1 long terminal repeat promoter by conversion of an NF- κ B enhancer element into a GABP binding site. *J Virol* 1999,**73**:1331-1340.
145. Biology LANLT, T- BG. *Human retroviruses and AIDS: a compilation and analysis of nucleic acid and amino acid sequences*: Theoretical Biology and Biophysics Group T-10, Los Alamos National Laboratory; 1998.
146. Commons W. HIV Virion. In; 2009.
147. Wilen CB, Tilton JC, Doms RW. HIV: cell binding and entry. *Cold Spring Harbor perspectives in medicine* 2012:a006866.
148. 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. 1984.
149. Clapham PR, McKnight Á. Cell surface receptors, virus entry and tropism of primate lentiviruses. *Journal of General Virology* 2002,**83**:1809-1829.

150. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhardt M, *et al.*
Identification of a major co-receptor for primary isolates of HIV-1. 1996.
151. Kozak SL, Platt EJ, Madani N, Ferro F, Peden K, Kabat D. CD4, CXCR-4,
and CCR-5 dependencies for infections by primary patient and laboratory-
adapted isolates of human immunodeficiency virus type 1. *J Virol*
1997,**71**:873-882.
152. Chiu TK, Davies DR. Structure and function of HIV-1 integrase. *Current*
topics in medicinal chemistry 2004,**4**:965-977.
153. Engelman A, Cherepanov P. The lentiviral integrase binding protein
LEDGF/p75 and HIV-1 replication. *PLoS pathogens* 2008,**4**:e1000046.
154. Fun A, Wensing A, Verheyen J, Nijhuis M. Human immunodeficiency virus
Gag and protease: partners in resistance. *Retrovirology* 2012,**9**:1-14.
155. Hornak V, Okur A, Rizzo RC, Simmerling C. HIV-1 protease flaps
spontaneously open and reclose in molecular dynamics simulations. *Proc Natl*
Acad Sci U S A 2006,**103**:915-920.
156. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE.
Characterization of ribosomal frameshifting in HIV-1 gag-pol expression.
Nature 1988,**331**:280-283.
157. Pettit SC, Everitt LE, Choudhury S, Dunn BM, Kaplan AH. Initial cleavage of
the human immunodeficiency virus type 1 GagPol precursor by its activated
protease occurs by an intramolecular mechanism. *J Virol* 2004,**78**:8477-8485.
158. Tang C, Louis JM, Aniana A, Suh J-Y, Clore GM. Visualizing transient events
in amino-terminal autoprocessing of HIV-1 protease. *Nature* 2008,**455**:693-
696.

159. de Oliveira T, Engelbrecht S, Janse van Rensburg E, Gordon M, Bishop K, zur Megede J, *et al.* Variability at Human Immunodeficiency Virus Type 1 Subtype C Protease Cleavage Sites: an Indication of Viral Fitness? *J Virol* 2003,**77**:9422-9430.
160. Das K, Martinez SE, Bauman JD, Arnold E. HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat Struct Mol Biol* 2012,**19**:253-259.
161. 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**:1669-1675.
162. Hsu P, Revolution V. Mechanism of HIV-1 Acquired Resistance to Protease Inhibitors and Its Possible Limitations.
163. Huff JR. HIV protease: a novel chemotherapeutic target for AIDS. *Journal of Medicinal Chemistry* 1991,**34**:2305-2314.
164. 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. *J Virol* 2003,**77**:4836-4847.
165. Administration USFaD. HIV/AIDS Treatment. In; 2014.
166. Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. *Cold Spring Harbor perspectives in medicine* 2012,**2**:a007161.
167. Ambrose Z, Aiken C. HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology* 2014,**454**:371-379.
168. Clavel F, Hance AJ. HIV Drug Resistance. *New England Journal of Medicine* 2004,**350**:1023-1035.

169. Chiba M, Jin L, Neway W, Vacca JP, Tata JR, Chapman K, *et al.* P450 interaction with HIV protease inhibitors: relationship between metabolic stability, inhibitory potency, and P450 binding spectra. *Drug metabolism and disposition* 2001,**29**:1-3.
170. Kempf DJ, Marsh KC, Kumar G, Rodrigues AD, Denissen JF, McDonald E, *et al.* Pharmacokinetic enhancement of inhibitors of the human immunodeficiency virus protease by coadministration with ritonavir. *Antimicrobial agents and chemotherapy* 1997,**41**:654-660.
171. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 1995,**373**:123-126.
172. Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, *et al.* Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 1995,**373**:117-122.
173. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* 1996,**271**:1582-1586.
174. Piatak M, Saag M, Yang L, Clark S, Kappes J, Luk K, *et al.* High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* 1993,**259**:1749-1754.
175. Maldarelli F, Kearney M, Palmer S, Stephens R, Mican J, Polis MA, *et al.* HIV Populations Are Large and Accumulate High Genetic Diversity in a Nonlinear Fashion. *J Virol* 2013,**87**:10313-10323.
176. Kahn JO, Walker BD. Acute human immunodeficiency virus type 1 infection. *New England Journal of Medicine* 1998,**339**:33-39.

177. Havlir DV, Richman DD. Viral dynamics of HIV: implications for drug development and therapeutic strategies. *Annals of internal medicine* 1996,**124**:984-994.
178. 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. *J Virol* 2010,**84**:9864-9878.
179. Roberts J, Preston B, Johnston L, Soni A, Loeb L, Kunkel T. Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro. *Molecular and cellular biology* 1989,**9**:469-476.
180. Jern P, Russell RA, Pathak VK, Coffin JM. Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. *PLoS pathogens* 2009,**5**:e1000367.
181. Sarafianos SG, Das K, Clark AD, Jr., Ding J, Boyer PL, Hughes SH, *et al.* Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids. *Proc Natl Acad Sci U S A* 1999,**96**:10027-10032.
182. Johnson VA, Calvez V, Günthard HF, Paredes R, Pillay D, Shafer RW, *et al.* Special Contribution - Update of the Drug Resistance Mutations in HIV-1: March 2013. *International AIDS Society-USA 2013, Topics in Antiviral Medicine*.
183. Tu X, Das K, Han Q, Bauman JD, Clark AD, Hou X, *et al.* Structural basis of HIV-1 resistance to AZT by excision. *Nature Structural & Molecular Biology* 2010,**17**:1202-1209.
184. Kozal M. Review: Cross-Resistance Patterns Among HIV Protease Inhibitors. *AIDS patient care and STDs* 2004,**18**:199-208.

185. Mao Y. Dynamical Basis for Drug Resistance of HIV-1 Protease. *BMC Structural Biology* 2011,**11**:1472-6807.
186. Shafer RW. Rationale and Uses of a Public HIV Drug - Resistance Database. *Journal of Infectious Diseases* 2006,**194**:S51-S58.
187. Dam E, Lebel-Binay S, Rochas S, Thibaut L, Faudon JL, Thomas CM, *et al.* Synergistic inhibition of protease-inhibitor-resistant HIV type 1 by saquinavir in combination with atazanavir or lopinavir. *Antiviral therapy* 2007,**12**:371.
188. Kantor R, Katzenstein D. Polymorphism in HIV-1 non-subtype B protease and reverse transcriptase and its potential impact on drug susceptibility and drug resistance evolution. *AIDS rev* 2003,**5**:25-35.
189. Kantor R, Katzenstein DA, Efron B, Carvalho AP, Wynhoven B, Cane P, *et al.* Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. *PLoS Medicine* 2005,**2**:e112.
190. Kantor R, Smeaton L, Vardhanabhuti S, Hudelson SE, Wallis CL, Tripathy S, *et al.* Pre-Treatment HIV Drug Resistance and HIV-1 Subtype C are Independently Associated with Virologic Failure: Results from the Multi-national PEARLS (ACTG A5175) Clinical Trial. *Clinical infectious diseases* 2015:civ102.
191. Lessells R, Katzenstein D, De Oliveira T. Are subtype differences important in HIV drug resistance? *Current opinion in virology* 2012,**2**:636-643.
192. Wainberg MA, Brenner BG. Role of HIV subtype diversity in the development of resistance to antiviral drugs. *Viruses* 2010,**2**:2493-2508.

193. Scherrer AU, Ledergerber B, von Wyl V, Böni J, Yerly S, Klimkait T, *et al.* Improved virological outcome in white patients infected with HIV-1 non-B subtypes compared to subtype B. *Clinical infectious diseases* 2011;cir669.
194. Touloumi G, Pantazis N, Chaix M-L, Bucher HC, Zangerle R, Kran A-MB, *et al.* Virologic and Immunologic Response to cART by HIV-1 Subtype in the CASCADE Collaboration. *PLoS ONE* 2013;**8**:e71174.
195. Amornkul PN, Karita E, Kamali A, Rida WN, Sanders EJ, Lakhi S, *et al.* Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa. *AIDS (London, England)* 2013;**27**:2775.
196. Gupta R, Hill A, Sawyer AW, Pillay D. Emergence of drug resistance in HIV type 1-infected patients after receipt of first-line highly active antiretroviral therapy: a systematic review of clinical trials. *Clinical infectious diseases* 2008;**47**:712-722.
197. Huang A, Katzenstein D, Saravanan S, Wu Y, Sirivichayakul S, Kumarasamy N, *et al.* Differential first-line antiretroviral therapy resistance in HIV-1 non-B subtypes in India, China and Thailand. In: *Antiviral therapy*: INT MEDICAL PRESS LTD 2-4 IDOL LANE, LONDON EC3R 5DD, ENGLAND; 2013. pp. A75-A75.
198. Asahchop EL, Oliveira M, Wainberg MA, Brenner BG, Moisi D, Toni TdA, *et al.* Characterization of the E138K resistance mutation in HIV-1 reverse transcriptase conferring susceptibility to etravirine in B and non-B HIV-1 subtypes. *Antimicrobial agents and chemotherapy* 2011;**55**:600-607.
199. Lai M-T, Lu M, Felock PJ, Hrin RC, Wang Y-J, Yan Y, *et al.* Distinct mutation pathways of non-subtype B HIV-1 during in vitro resistance

- selection with nonnucleoside reverse transcriptase inhibitors.
Antimicrobial agents and chemotherapy 2010,**54**:4812-4824.
200. Skhosana L, Steegen K, Bronze M, Lukhwareni A, Letsoalo E, Papathanasopoulos MA, *et al.* High Prevalence of the K65R Mutation in HIV-1 Subtype C Infected Patients Failing Tenofovir-Based First-Line Regimens in South Africa. *PLoS ONE* 2015,**10**:e0118145.
201. Sunpath H, Wu B, Gordon M, Hampton J, Johnson B, Moosa MY, *et al.* High rate of K65R for antiretroviral therapy-naive patients with subtype C HIV infection failing a tenofovir-containing first-line regimen. *Aids* 2012,**26**:1679-1684.
202. Martinez-Cajas JL, Wainberg MA, Oliveira M, Asahchop EL, Doualla-Bell F, Lisovsky I, *et al.* The role of polymorphisms at position 89 in the HIV-1 protease gene in the development of drug resistance to HIV-1 protease inhibitors. *Journal of Antimicrobial Chemotherapy* 2012:dkr582.
203. 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. *J Virol* 2009,**83**:9094-9101.
204. Parry CM, Kolli M, Myers RE, Cane PA, Schiffer C, Pillay D. Three residues in HIV-1 matrix contribute to protease inhibitor susceptibility and replication capacity. *Antimicrobial agents and chemotherapy* 2011,**55**:1106-1113.
205. Sutherland KA, Mbisa JL, Ghosn J, Chaix M-L, Cohen-Codar I, Hue S, *et al.* Phenotypic characterization of virological failure following lopinavir/ritonavir monotherapy using full-length gag–protease genes. *Journal of Antimicrobial Chemotherapy* 2014:dku296.

206. 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**:e1000345.
207. 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. *J Virol* 1996,**70**:3763-3769.
208. 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**:879.
209. Mepham S, Bland R, Newell ML. Prevention of mother - to - child transmission of HIV in resource - rich and - poor settings. *BJOG: An International Journal of Obstetrics & Gynaecology* 2011,**118**:202-218.
210. C K, B F, T L, C A, B H, I M, *et al.* HIV Sequence Compendium 2010. *Theoretical Biology and Biophysics Group, Los Alamos National Laboratory* 2010.
211. Butler DM, Pacold ME, Jordan PS, Richman DD, Smith DM. The efficiency of single genome amplification and sequencing is improved by quantitation and use of a bioinformatics tool. *Journal of virological methods* 2009,**162**:280-283.
212. Mbisa JL, Gupta RK, Kabamba D, Mulenga V, Kalumbi M, Chintu C, *et al.* The evolution of HIV-1 reverse transcriptase in route to acquisition of Q151M

- multi-drug resistance is complex and involves mutations in multiple domains. *Retrovirology* 2011,**8**:31.
213. Silvestro D, Michalak I. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution* 2012,**12**:335-337.
214. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution* 1981,**17**:368-376.
215. Felsenstein J. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Biology* 1973,**22**:240-249.
216. Salemi M, Vandamme A-M, Lemey P. *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*: Cambridge University Press; 2009.
217. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* 2011,**28**:2731-2739.
218. Pond SLK, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and evolution* 2005,**22**:1208-1222.
219. Kosakovsky Pond S, Poon A, Zarate S, Smith D, Little S, Pillai S, *et al.* Estimating selection pressures on HIV - 1 using phylogenetic likelihood models. *Statistics in medicine* 2008,**27**:4779-4789.
220. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Pond SLK, *et al.* FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular biology and evolution* 2013:mst030.

221. Delpont W, Poon AF, Frost SD, Pond SLK. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 2010,**26**:2455-2457.
222. Kearney MF, Spindler J, Shao W, Yu S, Anderson EM, O'Shea A, *et al.* Lack of Detectable HIV-1 Molecular Evolution during Suppressive Antiretroviral Therapy. *PLoS pathogens* 2014,**10**:e1004010.
223. Riva C, Violin M, Cozzi-Lepri A, Velleca R, Bertoli A, Vigano P, *et al.* Transmitted virus with substitutions at position 215 and risk of virological failure in antiretroviral-naïve patients starting highly active antiretroviral therapy. *Antivir Ther* 2002,**7**:S136-S136.
224. Violin M, Cozzi-Lepri A, Velleca R, Vincenti A, D'Elia S, Chiodo F, *et al.* Risk of failure in patients with 215 HIV-1 revertants starting their first thymidine analog-containing highly active antiretroviral therapy. *Aids* 2004,**18**:227-235.
225. Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science* 1988,**242**:1171-1173.
226. John C Hall, Brian J Hall, Cockerel CJ. Laboratory Testing for HIV/AIDS. In: *HIV/AIDS in the Post-HAART Era: Manifestations, Treatment, Epidemiology*. Edited by Tang Y-W: McGraw-Hill Medical; 1 Har/Dgd edition (1 Sep 2011); 2011. pp. 128-129.
227. GÜNTARD HF, WONG JK, IGNACIO CC, HAVLIR DV, RICHMAN DD. Comparative performance of high-density oligonucleotide sequencing and dideoxynucleotide sequencing of HIV type 1 pol from clinical samples. *AIDS Res Hum Retroviruses* 1998,**14**:869-876.

228. Shafer RW, Winters MA, Palmer S, Merigan TC. Multiple concurrent reverse transcriptase and protease mutations and multidrug resistance of HIV-1 isolates from heavily treated patients. *Annals of internal medicine* 1998,**128**:906-911.
229. Ross L, Vavro C, Wine B, Liao Q. Fosamprenavir clinical study meta-analysis in ART-naive subjects: rare occurrence of virologic failure and selection of protease-associated mutations. *HIV clinical trials* 2006,**7**:334-338.
230. Malan DN, Krantz E, David N, Wirtz V, Hammond J, McGrath D, *et al.* Efficacy and safety of atazanavir, with or without ritonavir, as part of once-daily highly active antiretroviral therapy regimens in antiretroviral-naive patients. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2008,**47**:161-167.
231. Benaboud S, Ekouevi DK, Urien S, Rey E, Arrive E, Blanche S, *et al.* Population pharmacokinetics of nevirapine in HIV-1-infected pregnant women and their neonates. *Antimicrob Agents Chemother* 2011,**55**:331-337.
232. Johnson VA, Petropoulos CJ, Woods CR, Hazelwood JD, Parkin NT, Hamilton CD, *et al.* Vertical transmission of multidrug-resistant human immunodeficiency virus type 1 (HIV-1) and continued evolution of drug resistance in an HIV-1–infected infant. *Journal of Infectious Diseases* 2001,**183**:1688-1693.
233. van Zyl GU, van der Merwe L, Claassen M, Cotton MF, Rabie H, Prozesky HW, *et al.* Protease inhibitor resistance in South African children with virologic failure. *The Pediatric Infectious Disease Journal* 2009,**28**:1125-1127.

234. 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**:760-766.
235. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014,**30**:1312-1313.
236. Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular biology and evolution* 2006,**23**:1891-1901.
237. Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SD. GARD: a genetic algorithm for recombination detection. *Bioinformatics* 2006,**22**:3096-3098.
238. Pond SLK, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 2005,**21**:2531-2533.
239. Kieffer TL, Finucane MM, Nettles RE, Quinn TC, Broman KW, Ray SC, *et al.* Genotypic analysis of HIV-1 drug resistance at the limit of detection: virus production without evolution in treated adults with undetectable HIV loads. *Journal of Infectious Diseases* 2004,**189**:1452-1465.
240. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, *et al.* Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* 1999,**73**:10489-10502.
241. Mullen J, Leech S, O'Shea S, Chrystie IL, du Mont G, Ball C, *et al.* Antiretroviral drug resistance among HIV - 1 infected children failing treatment. *Journal of medical virology* 2002,**68**:299-304.
242. Lange J, Hill A, Ait-Khaled M, Foudraire N, Maguire M. Does the M184V mutation affect the efficacy of HAART? Evidence from sequential addition of

- protease inhibitors or abacavir to zidovudine/lamivudine or stavudine/lamivudine. In: *Third International Workshop on HIV Drug Resistance and Treatment Strategies*. San Diego, USA; 1999.
243. Permanyer M, Ballana E, Ruiz A, Badia R, Riveira-Munoz E, Gonzalo E, *et al.* Antiretroviral agents effectively block HIV replication after cell-to-cell transfer. *J Virol* 2012,**86**:8773-8780.
244. Sigal A, Kim JT, Balazs AB, Dekel E, Mayo A, Milo R, *et al.* Cell-to-cell spread of HIV permits ongoing replication despite antiretroviral therapy. *Nature* 2011,**477**:95-98.
245. Poon AF, Lewis FI, Frost SD, Pond SLK. Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models. *Bioinformatics* 2008,**24**:1949-1950.
246. Poon AF, Lewis FI, Pond SLK, Frost SD. An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. *PLoS computational biology* 2007,**3**:e231.
247. Poon AF, Pond SLK, Richman DD, Frost SD. Mapping protease inhibitor resistance to human immunodeficiency virus type 1 sequence polymorphisms within patients. *J Virol* 2007,**81**:13598-13607.
248. Mulder LC, Harari A, Simon V. Cytidine deamination induced HIV-1 drug resistance. *Proceedings of the National Academy of Sciences* 2008,**105**:5501-5506.
249. Amoêdo ND, Afonso AO, Cunha SM, Oliveira RH, Machado ES, Soares MA. Expression of APOBEC3G/3F and G-to-A hypermutation levels in HIV-1-infected children with different profiles of disease progression. *PLoS ONE* 2011,**6**:e24118.

250. Koulinska IN, Chaplin B, Mwakagile D, Essex M, Renjifo B. Hypermutation of HIV type 1 genomes isolated from infants soon after vertical infection. *AIDS Res Hum Retroviruses* 2003,**19**:1115-1123.
251. Torrecilla E, Delicado TL, Holguín Á. New Findings in Cleavage Sites Variability across Groups, Subtypes and Recombinants of Human Immunodeficiency Virus Type 1. *PLoS ONE* 2014,**9**:e88099.
252. Collins JA, Thompson MG, Paintsil E, Ricketts M, Gedzior J, Alexander L. Competitive Fitness of Nevirapine-Resistant Human Immunodeficiency Virus Type 1 Mutants. *J Virol* 2003,**78**:603-611.
253. Iglesias-Ussel MD, Casado C, Yuste Es, Olivares I, López-Galíndez C. In vitro analysis of human immunodeficiency virus type 1 resistance to nevirapine and fitness determination of resistant variants. *Journal of General Virology* 2002,**83**:93-101.
254. Hu Z, Kuritzkes DR. Altered viral fitness and drug susceptibility in human immunodeficiency virus type 1 (HIV-1) carrying mutations that confer resistance to non-nucleoside reverse transcriptase and integrase strand-transfer inhibitors. *J Virol* 2014:JVI. 00695-00614.
255. Hu Z, Kuritzkes DR. Interaction of reverse transcriptase (RT) mutations conferring resistance to lamivudine and etravirine: effects on fitness and RT activity of human immunodeficiency virus type 1. *J Virol* 2011,**85**:11309-11314.
256. Miller M, White K, Petropoulos C, Parkin N. Decreased replication capacity of HIV-1 clinical isolates containing K65R or M184V RT mutations. In: *10th Conference on Retroviruses and Opportunistic Infections (CROI), Boston, 10–14 February 2003*; 2003.

257. Mo H, King MS, King K, Molla A, Brun S, Kempf DJ. Selection of resistance in protease inhibitor-experienced, human immunodeficiency virus type 1-infected subjects failing lopinavir-and ritonavir-based therapy: mutation patterns and baseline correlates. *J Virol* 2005,**79**:3329-3338.
258. 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. *J Virol* 2001,**75**:7462-7469.
259. Bessong PO. Polymorphisms in HIV - 1 subtype C proteases and the potential impact on protease inhibitors. *Tropical Medicine & International Health* 2008,**13**:144-151.
260. Holguín A, Paxinos E, Hertogs K, Womac C, Soriano V. Impact of frequent natural polymorphisms at the protease gene on the in vitro susceptibility to protease inhibitors in HIV-1 non-B subtypes. *Journal of clinical virology* 2004,**31**:215-220.
261. Abecasis AB, Deforche K, Bachelier LT, McKenna P, Carvalho AP, Gomes P, *et al.* Investigation of baseline susceptibility to protease inhibitors in HIV-1 subtypes C, F, G and CRF02_AG. *Antiviral therapy* 2006,**11**:581.
262. Velázquez-Campoy A, Vega S, Fleming E, Bacha U, Sayed Y, Dirr HW, *et al.* Protease inhibition in African subtypes of HIV-1. *AIDS rev* 2003,**5**:165-171.
263. Velazquez-Campoy A, Vega S, Freire E. Amplification of the Effects of Drug Resistance Mutations by Background Polymorphisms in HIV-1 Protease from African Subtypes†. *Biochemistry* 2002,**41**:8613-8619.

264. Winters MA, Bosch RJ, Albrecht MA. Clinical impact of the M184V mutation on switching to didanosine or maintaining lamivudine treatment in nucleoside reverse-transcriptase inhibitor-experienced patients. *Journal of Infectious Diseases* 2003,**188**:537-540.
265. Petrella M, Oliveira M, Moisi D, Detorio M, Brenner BG, Wainberg MA. Differential maintenance of the M184V substitution in the reverse transcriptase of human immunodeficiency virus type 1 by various nucleoside antiretroviral agents in tissue culture. *Antimicrobial agents and chemotherapy* 2004,**48**:4189-4194.
266. Maguire MF, Guinea R, Griffin P, Macmanus S, Elston RC, Wolfram J, *et al.* Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *J Virol* 2002,**76**:7398-7406.
267. Kolli M, Stawiski E, Chappey C, Schiffer CA. Human immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance inhibitor resistance. *J Virol* 2009,**83**:11027-11042.
268. Clavel F, Mammano F. Role of Gag in HIV resistance to protease inhibitors. *Viruses* 2010,**2**:1411-1426.
269. Novitsky V, Wang R, Margolin L, Baca J, Kebaabetswe L, Rossenkhan R, *et al.* Timing constraints of in vivo gag mutations during primary HIV-1 subtype C infection. *PLoS ONE* 2009,**4**:e7727.
270. Novitsky V, Wang R, Margolin L, Baca J, Moyo S, Musonda R, *et al.* Dynamics and timing of invivo mutations at Gag residue 242 during primary HIV-1 subtype C infection. *Virology* 2010,**403**:37-46.

271. Barth RE, Tempelman HA, Smelt E, Wensing AM, Hoepelman AI, Geelen SP. Long-term outcome of children receiving antiretroviral treatment in rural South Africa: substantial virologic failure on first-line treatment. *The Pediatric Infectious Disease Journal* 2011,**30**:52-56.
272. Eley BS, Meyers T. Antiretroviral Therapy for Children in Resource-Limited Settings. *Pediatric Drugs* 2011,**13**:303-316.
273. Shuter J, Sarlo JA, Kanmaz TJ, Rode RA, Zingman BS. HIV-infected patients receiving lopinavir/ritonavir-based antiretroviral therapy achieve high rates of virologic suppression despite adherence rates less than 95%. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2007,**45**:4-8.
274. Moyle G. Metabolic issues associated with protease inhibitors. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2007,**45**:S19-S26.
275. Ren Y, Nuttall JJ, Egbers C, Eley BS, Meyers TM, Smith PJ, *et al.* Effect of rifampicin on lopinavir pharmacokinetics in HIV-infected children with tuberculosis. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2008,**47**:566-569.
276. Ananworanich J, Kosalaraksa P, Siangphoe U, Engchanil C, Pancharoen C, Lumbiganon P, *et al.* A feasibility study of immediate versus deferred antiretroviral therapy in children with HIV infection. *AIDS Res Ther* 2008,**5**:24.
277. Puthanakit T, Saphonn V, Ananworanich J, Kosalaraksa P, Hansudewechakul R, Vibol U, *et al.* Early versus deferred antiretroviral therapy for children older than 1 year infected with HIV (PREDICT): a multicentre, randomised, open-label trial. *The Lancet infectious diseases* 2012,**12**:933-941.

278. Puthanakit T, Ananworanich J, Vonthanak S, Kosalaraksa P, Hansudewechakul R, van der Lugt J, *et al.* Cognitive function and neurodevelopmental outcomes in HIV-infected children older than 1 year of age randomized to early versus deferred antiretroviral therapy: the PREDICT neurodevelopmental study. *The Pediatric Infectious Disease Journal* 2013,**32**:501.
279. Department of Health RoSA. Guidelines for the management of tuberculosis in children. In. Edited by Health Do. Republic of South Africa: Department of Health, Private Bag X828, Pretoria 0001, South Africa; 2013.
280. La Porte C, Colbers E, Bertz R, Voncken D, Wikstrom K, Boeree M, *et al.* Pharmacokinetics of adjusted-dose lopinavir-ritonavir combined with rifampin in healthy volunteers. *Antimicrobial agents and chemotherapy* 2004,**48**:1553-1560.
281. McIlleron H, Ren Y, Nuttall J, Fairlie L, Rabie H, Cotton M, *et al.* Short communication Lopinavir exposure is insufficient in children given double doses of lopinavir/ritonavir during rifampicin-based treatment for tuberculosis. *Antiviral therapy* 2011,**16**:417-421.
282. Zhang C, McIlleron H, Ren Y, van der Walt J-S, Karlsson MO, Simonsson US, *et al.* Population pharmacokinetics of lopinavir and ritonavir in combination with rifampicin-based antitubercular treatment in HIV-infected children. *Antiviral therapy* 2012,**17**:25.
283. Hedt BL, Laufer MK, Cohen T. Drug resistance surveillance in resource-poor settings: current methods and considerations for TB, HIV, and malaria. *The American journal of tropical medicine and hygiene* 2011,**84**:192-199.

284. van Zyl GU, Cotton MF, Claassen M, Abrahams C, Preiser W. Surveillance of transmitted resistance to antiretroviral drug classes among young children in the Western Cape Province of South Africa. *The Pediatric Infectious Disease Journal* 2010,**29**:370-371.
285. 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-Naïve Populations and Associate with Reduced Treatment Efficacy. *PLoS Med* 2008,**5**:e158.
286. Stringer JSA, McConnell MS, Kiarie J, Bolu O, Anekthananon T, Jariyasethpong T, *et al.* Effectiveness of Non-nucleoside Reverse-Transcriptase Inhibitor-Based Antiretroviral Therapy in Women Previously Exposed to a Single Intrapartum Dose of Nevirapine: A Multi-country, Prospective Cohort Study. *PLoS Med* 2010,**7**:e1000233.
287. Babiker A, Compagnucci A, Fiscus S, Giaquinto C, Gibb D, Harper L, *et al.* First-line antiretroviral therapy with a protease inhibitor versus non-nucleoside reverse transcriptase inhibitor and switch at higher versus low viral load in HIV-infected children: an open-label, randomised phase 2/3 trial. *The Lancet infectious diseases* 2011,**11**:273-283.
288. Persaud D, Palumbo P, Ziemniak C, Chen J, Ray SC, Hughes M, *et al.* Early archiving and predominance of nonnucleoside reverse transcriptase inhibitor-resistant HIV-1 among recently infected infants born in the United States. *Journal of Infectious Diseases* 2007,**195**:1402-1410.
289. Wind-Rotolo M, Durand C, Cranmer L, Reid A, Martinson N, Doherty M, *et al.* Identification of nevirapine-resistant HIV-1 in the latent reservoir after

single-dose nevirapine to prevent mother-to-child transmission of HIV-1.

Journal of Infectious Diseases 2009,**199**:1301-1309.

290. Scherrer AU, Ledergerber B, von Wyl V, Böni J, Yerly S, Klimkait T, *et al.* Minor protease inhibitor mutations at baseline do not increase the risk for a virological failure in HIV-1 subtype B infected patients. *PLoS ONE* 2012,**7**:e37983.
291. Ong LY, Razak SNH, Lee YM, Sri La Sri Ponnampalavanar S, Syed Omar SF, Azwa RI, *et al.* Molecular diversity of HIV - 1 and surveillance of transmitted drug resistance variants among treatment Naïve patients, 5 years after active introduction of HAART in Kuala Lumpur, Malaysia. *Journal of medical virology* 2014,**86**:38-44.
292. Control CfD. Managing Drug Interactions in the Treatment of HIV-related Tuberculosis. In. Edited by Elimination DoT. Atlanta, GA, USA 30333; 2013.