Minority Drug Resistant Variants of HIV-1 and Response to Early Combination Therapy

Jessica R. Keys

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology in the Gillings School of Global Public Health.

Chapel Hill
2014

Approved by:

Charles Poole

Sonia Napravnik

Joseph J. Eron, Jr.

Ronald Swanstrom

Steven Meshnick

© 2014
Jessica R. Keys
ALL RIGHTS RESERVED

ABSTRACT

JESSICA R. KEYS: Minority Drug Resistant Variants of HIV-1 and Response to Early Combination Therapy
(Under the direction of Charles Poole)

Initial HIV-1 therapy selection is informed by sequencing of a bulk PCR product to screen for antiretroviral resistance mutations. However, this method does not reliably sample drug resistant variants that occur in <20% of the viral population, and these may re-emerge and impair treatment response once therapy is administered. Alternatively, ultra deep sequencing can detect minority drug resistant variants, but it is difficult to distinguish very low abundance mutations from error. To address deep sequencing error, two regions of the HIV-1 genome spanning reverse transcriptase (RT) codons 34-245 were tagged with a random 8-nucleotide sequence (Primer ID) prior to PCR and sequencing. Primer ID allowed us to use resampled raw sequences sharing the same Primer ID to construct consensus sequences, each representing an original viral template within that sample.

We first established a residual error rate for Primer ID using known sequences for both the Roche 454 and Illumina MiSeq deep sequencing platforms. Primer ID reduced 454 and MiSeq errors from 71 to 2.6 and from 24 to 1.2 errors/10,000 nucleotides, respectively. Applying Primer ID corrected 454 deep sequencing to 184 therapy-naïve patients from North Carolina that went on to receive RT inhibitor based combination therapy, we found that 14% of had at least one RT inhibitor mutation, compared to 2.7% using standard bulk sequence analysis. Nearly 10% of 184 patients received regimens that contained fewer than 3 active antiretrovirals, according to the

Stanford resistance algorithm. While patients on suboptimal therapy failed faster than patients on fully-active regimens, the effect was driven by resistance detected by standard methods rather than previously undetected minority variants. Overall, the use of Primer ID revealed limited template utilization, limiting the depth of deep sequencing sampling. Primer ID addresses important limitations of deep sequencing and produces less biased estimates of low level resistance mutations in the viral population, which may allow us to more accurately define a threshold at which minority drug resistant variants of HIV-1 begin to compromise treatment response.

ACKNOWLEDGEMENTS

None of this work would have been possible without the help of my committee members, past and present included. First, I would like to thank my chair, Charlie Poole, who has made sure that I remember the fundamentals of epidemiology in almost every context imaginable. He has helped me organize my thoughts in a logical manner, and he has been supportive of my project from the beginning. I would secondly like to thank Ron Swanstrom, who provided not only his knowledge, but also a home in his lab for the last six years. I may not be a card-carrying Virologist, but Ron has allowed me a rare opportunity to pursue my overlapping interests in epidemiology and molecular biology. Speaking of overlap, I would also like to thank Joe Eron, who has always tied everything together for me, even when it seemed that everyone was speaking a different language. He has a talent for seeing the big picture and a deep understanding of many disciplines, and I benefitted greatly from his insights. I would also like to thank Sonia Napravnik for her expert advice. Sonia has been approachable from day one and an invaluable asset in developing my dissertation topic. I would also like to thank Steve Meshnick, who came in towards the tail end of my dissertation and was essential in helping me complete this work. I thank him for reviewing my papers and keeping me on task. I would like to thank a former committee member, Jeff Anderson, who has since moved on from UNC. Without Jeff, I would have never have figured out the protocols for deep sequencing. Jeff was always in the lab, and would happily trade his valuable time for a Snickers bar.

I would also like to acknowledge the financial support I have received over the years. Specifically, I am grateful for the funding I received from the UNC-NIH AIDS and

STD Training Program, which allowed me to travel abroad and present my work. I would like to thank the UNC Center for AIDS Research for funding my supplemental projects, allowing me to continue my dissertation work while simultaneously learning about other research avenues. I would like to thank the NC Translational and Clinical Sciences Institute for providing the funding for deep sequencing through their Pilot grant program.

Of course, I am sincerely grateful to my friends and family. My parents have provided me with their support throughout this long process, and both have drawn on their own experiences in academia to help me succeed. I am also thankful for my contemporaries in epidemiology, including the members of the former Dissertation Support Group. Lizzi Torrone, Sandi McCoy, Brooke Levandowski, Brooke Hoots, Kelly Quinn, Kim Porter, and Kim Powers have all moved on in their careers, but I will never forget the moral and scholastic support that these young women provided me. I would also like to thank members of the Swanstrom lab for their help at the bench, especially Suani Ojeda and Li Hua Ping for their help in troubleshooting my experiments, and Shuntai Zhou for processing all of the raw deep sequencing data. Finally, I want to thank Carmen Woody and Nancy Colvin of the Epidemiology Department, who have encouraged me along the way and helped me navigate the bureaucratic process.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xi

LIST OF ABBREVIATIONS xii

Chapter 1 SPECIFIC AIMS 1

1.1 Statement of Purpose 1

1.2 Specific Aim 1 4

1.3 Specific Aim 2 5

1.4 Specific Aim 3 6

Chapter 2 BACKGROUND 7

2.1 Epidemiology of HIV Infection 7

2.2 HIV Virology 9

2.3 Natural History of HIV Infection 11

2.4 Treatment Strategies for HIV Infection 15

2.5 Selection of Resistance Mutations with Antiretroviral Therapy 21

2.6 Mechanisms of Reverse Transcriptase Inhibitor Resistance 23

2.7 Epidemiology of Acquired Antiretroviral Resistance 25

2.8 Epidemiology of Transmitted Reverse Transcriptase Inhibitor Resistance 30

2.9 Measurement of Antiretroviral Associated Resistance 45

Chapter 3 RESEARCH DESIGN AND METHODS 57

3.1 Overview 57

3.2 Study Population 58

3.3 The University of North Carolina Center for AIDS Research 60

3.4 Data Sources 61

3.5 HIV-Related Laboratory Procedures 62

3.6 Statistical Analyses 71

3.7 Limitations 90

Chapter 4 PRIMER ID CORRECTS NEXT-GENERATION SEQUENCING PLATFORMS AND STILL REVEALS PRE-EXISTING DRUG RESISTANCE MUTATIONS IN THE HIV-1 REVERSE TRANSCRIPTASE CODING DOMAIN 94

4.1 Introduction 94

4.2 Methods 95

4.3 Results 100

4.4 Discussion 112

Chapter 5 VIROLOGIC CONSEQUENSES OF PRE-EXISTING LOW ABUNDANCE RESISTANCE AMONG PATIENTS INITIATING WITH A REVERSE TRANSCRIPTASE INHIBITOR REGIMEN 118

5.1 Introduction 118

5.2 Methods 119

5.3 Results 122

5.4 Discussion 126

Chapter 6 CONCLUSION 129

6.1 Summary of Results 129

6.2 Implications 134

6.3 Future Directions 136

Appendix 2.1 MUTATIONS ASSOCIATED WITH REVERSE TRANSCRIPTASE INHIBITOR RESISTANCE, INTERANTIONAL AIDS SOCIETY-USA, 2013 [183] 138

Appendix 2.2 SURVEILLANCE DRUG RESISTANCE MUTATIONS FOR EVIDENCE OF TRANSMITTED RESISTANCE, 2009 [189] 140

Appendix 2.3 STANFORD HIV DRUG RESISTANCE DATABASE ANTIRETROVIRAL SUSCEPTIBILITY SCORES FOR HIV-1 REVERSE TRANSCRIPTASE MUTATIONS [171] 141

Appendix 3.1 PRIMERS FOR cDNA SYNTHESIS AND DNA AMPLIFICATION 144

Appendix 4.1 ERROR RATES ESTIMATED FROM DEEP SEQUENCING CONTROLS 145

Appendix 5.1 ADJUSTED HAZARD RATIOS USING STRICTER INTERPRETATION OF DEEP SEQUENCING DATA....146

REFERENCES....147

LIST OF TABLES

TABLE 2.1 Reverse transcriptase resistance among antiretroviral-naïve individuals with recent HIV-1 infection 34

TABLE 2.2 Reverse transcriptase inhibitor resistance among antiretroviral-naïve individuals with HIV-1 infection of unknown duration....................36

TABLE 2.3 Common genotypic assays to detect low abundance antiretroviral resistant HIV-1....................53

TABLE 3.1 Variables and specimens from the University of North Carolina Center for AIDS Research HIV Clinical Cohort Study 62

TABLE 4.1. Baseline characteristics of clinical participants....................105

TABLE 4.2. Resistance mutations within clinical participant samples estimated using raw or Primer ID sequences....................106

TABLE 5.1. Baseline characteristics of patient population by number of active drugs 123

TABLE 5.2. Adjusted hazard ratios for time-to-virologic failure and receiving suboptimal therapy 126

LIST OF FIGURES

FIGURE 2.1 AIDS cases, deaths, and persons living with AIDS, 1985-2009, US 8

FIGURE 2.2 Human immunodeficiency virus-1-encoded proteins........ 11

FIGURE 2.3 Classic model of HIV disease progression in the untreated HIV-infected adult 12

FIGURE 2.4 Antiretrovirals approved by the US Food and Drug Administration for the treatment of HIV-1 infection, 1987-2011 16

FIGURE 2.5 Primer ID method to estimate HIV-1 viral population 56

FIGURE 3.1 Flow diagram of the study population 60

FIGURE 3.2 Sampling depth and power to detect minority HIV-1 sequence variants........ 75

FIGURE 3.3 Prevalence and 95% confidence intervals by sample size 80

FIGURE 3.4 Directed acyclic graph of the causal association between pre-therapy resistance and time-to-virologic failure........ 88

FIGURE 3.5 Statistical power to detect various hazard ratios 90

FIGURE 4.1 Primer ID to estimate the HIV-1 population........ 97

FIGURE 4.2 Percent incorrect nucleotide calls for each queried position of control sequence 103

FIGURE 4.3 Participants with pre-existing RT inhibitor resistance mutations........ 108

FIGURE 4.4 Prevalence and relative abundance of pre-existing RT inhibitor resistance among clinical participants........ 110

FIGURE 5.1. Activity of early therapy predicted using Primer ID corrected deep sequencing genotype and time-to-virologic failure 125

LIST OF ABBREVIATIONS

AIDS	acquired immunodeficiency syndrome
ART	antiretroviral therapy
ARV	antiretroviral
GART	genotypic antiretroviral resistance test
HAART	highly active antiretroviral therapy
HIV	human immunodeficiency virus
IDU	injection drug user
MDR	multidrug resistance
MSM	men who have sex with men
NNRTI	non-nucleoside reverse transcriptase inhibitor
NRTI	nucleoside reverse transcriptase inhibitor
PI	protease inhibitor
PR	protease
RT	reverse transcriptase
RTI	reverse transcriptase inhibitor
SDRM	surveillance drug resistance mutation
TAM	thymidine analogue mutation
TDR	transmitted drug resistance
UDS	ultra deep sequencing

CHAPTER 1 SPECIFIC AIMS

1.1 STATEMENT OF PURPOSE

HIV-related morbidity and mortality have declined significantly since the introduction of highly active antiretroviral therapy (HAART) [1, 2]. However, the benefits of potent combination antiretroviral therapy may be offset by the development of resistance and cross-resistance to these drugs [3]. Furthermore, HIV-1 variants with antiretroviral (ARV) resistance mutations may be transmitted to others, possibly limiting efficacy of HAART among patients initiating antiretroviral therapy [4, 5].

US Department of Health and Human Services (DHHS) Guidelines recommend resistance testing combined with expert opinion in patients failing therapy [6]. However, DHHS recommendations have only recently favored testing *all* patients prior to initiating therapy, irrespective of duration of HIV infection. DHHS's caution stems from the fact that commercially-available genotypic antiretroviral resistance tests (GART) are unable to reliably detect minority HIV-1 variants present in below 20% of the population [7, 8]. Low sensitivity for minority variants could be problematic for testing chronically infected therapy-naïve patients, since resistant variants may have been overgrown by HIV-1 variants with better replicative capacity in the absence of drug pressure. However, once corresponding drugs are administered, it is presumed that these minority drug resistant variants would quickly re-emerge [9], warranting further exploration into the prevalence, diversity, and clinical impact of minority resistance populations [10].

The clinical utility of resistance testing in therapy-naïve patients is complex and depends on some key factors: transmissibility of resistance mutations; prevalence of

resistance in treated and untreated populations; persistence of resistance over time without drug pressure; and clinical relevance of minority drug-resistant HIV variants. Prevalence estimated using standard bulk sequence analysis indicates the burden of pre-therapy drug resistance is substantial in populations where ARV drugs have been historically more available, with an estimated 10-15% of untreated individuals having evidence of resistance [4, 5, 11, 12]. Moreover, transmitted HIV-1 variants with resistance mutations may persist as major circulating viral populations in therapy-naïve individuals for long periods of time [13-16]. While it is generally accepted that dominant resistant variants lead to adverse clinical outcomes in patients initiating therapy [5, 17], the impact of minority resistant variants on clinical outcomes is not as well characterized.

Recent studies have implicated minority non-nucleoside reverse transcriptase inhibitor (NNRTI) resistant variants in shorter time to virologic rebound [18], but a threshold at which resistance mutations are clinically relevant has not been clearly defined. Difficulty in defining a clinical cut-point for variants with resistance mutations is partly due to complications associated with their measurement. Specifically, rare HIV-1 RNA must be enriched in patient samples before sequencing, first by reverse transcription of HIV-1 RNA into cDNA followed by polymerase chain reaction (PCR) mediated amplification of the cDNA template. While both steps may introduce errors into the HIV sequence, PCR amplification in particular is associated with nucleotide misincorporation, recombination of viral sequences by template switching, biased amplification, and resampling of viral templates due to low template input [19, 20]. Further, HIV-1 replicates rapidly without a proofreading mechanism so that, theoretically, any mutation may be present in the viral population at any one time [21]. It is therefore critical to accurately estimate each individual’s viral population, which may be highly

variable and contain very low abundance resistance mutations, before a clinically relevant threshold can be defined.

Ultra deep sequencing (UDS) is a high throughput technique that promises sampling depths capable of detecting minority drug resistant variants of HIV-1 [22, 23]. For this collection of highly sensitive sequencing platforms, patient-derived HIV-1 nucleic acids are amplified, pooled, and sequenced in a massively parallel fashion. Sensitivity is mainly limited by the number of HIV-1 RNA templates input and the efficiency with which they are reverse transcribed and amplified. However, UDS, like all other methods that rely on HIV-1 amplification, is subject to errors associated with PCR. UDS platforms in which a single nucleotide is added at one time (e.g., the 454 sequencing platform) are additionally prone to nucleotide miscalls within nucleotide repeats (homopolymeric regions) [24].

Errors accumulated during UDS impede estimation of the highly diverse viral populations, since these must be distinguished from genuine mutations. Typically, known control sequences are used to inform statistical models that correct estimates skewed by nucleotide misincorporation, but not PCR re-sampling [25]. Alternatively, Jabara et al. introduced a method to address errors stemming from the PCR step forward, including PCR re-sampling [26]. With this approach, a random 8-nucleotide sequence is incorporated during cDNA synthesis so that each individual viral template is tagged with a unique Primer ID. After amplification and sequencing, majority-rules consensus sequences are constructed from sequences sharing the same Primer ID so that, collectively, these consensus sequences reflect the true viral population sampled rather than what was best amplified. Nucleotide miscalls are also filtered out by Primer ID consensus sequences since random nucleotide errors are more likely to be represented among the minority of raw sequences.

In this study, we estimated the prevalence of minority reverse transcriptase inhibitor (RTI)-resistant HIV-1 variants among therapy-naïve patients seeking HIV care at the University of North Carolina Infectious Disease Clinic and examined the influence of these resistant variants on response to early RTI-based HAART. For the latter question, we attempted to articulate relationships between the distribution of resistant variant within the individual and time-to-virologic failure (VF). To measure minority variants, we used an assay that is more sensitive than standard bulk sequence analysis to detect and quantify resistance in archived plasma samples from HIV-1-positive North Carolinians enrolled in the University of North Carolina Center for AIDS Research HIV Clinical Cohort (UCHCC). This assay was based on analysis of ultra deep sequencing (UDS) data generated by the 454 sequencing platform with the Primer ID method to allow more accurate estimation of the viral population [26]. We first established a baseline error rate for deep sequencing using both the 454 and less homopolymeric error prone Illumina platforms in a series of control experiments (Aim 1). We compared the results of resistance testing by deep sequencing with the 454 platform and standard bulk sequencing in samples from 184 patients obtained prior to therapy initiation (Aims 2A and 2B). Among a subset of 19 patients with the greatest number of Primer ID consensus sequences generated using the 454 platform, we compared results obtained using the 454 platform to those obtained using the Illumina MiSeq platform, which is not susceptible to homopolymeric errors (Aim 2C). Last, we estimated the association of pre-therapy RT inhibitor resistant HIV-1 variants with virologic response to early RT inhibitor-based HAART (Aim 3).

1.2 SPECIFIC AIM 1

Using a known HIV-1 reverse transcriptase sequence, define a baseline error rate for the 454 and MiSeq deep sequencing platforms. For each platform, compare results

estimated using Primer ID consensus sequences to those estimated using raw sequences.

Hypothesis 1.1: Using Primer ID to create consensus sequences will reduce PCR and sequencing errors compared to raw sequences.

Hypothesis 1.2: Error rates estimated using data from the 454 sequencing platform will be higher in homopolymeric tracts compared to stretches of heterogeneous nucleotides.

1.3 Specific Aim 2

(A) Using the 454 FLX platform to sequence HIV-1 RT codons 34-138 and 149-236, estimate the prevalence and relative abundance of RT inhibitor resistance among 184 therapy-naïve patients with concurrent bulk sequencing results. (B) Compare estimates obtained using the 454 platform to estimates obtained using bulk sequencing. (C) Among a subset of 19 patients with the greatest depth of sampling of viral templates, compare mutations detected within HIV-1 RT codons 34-73 and 111-138 between the Illumina MiSeq and Roche 454 platforms.

Hypothesis 2A: Some drug resistance mutations will be detected more often as minority variants, particularly those associated with a reduction in fitness or those that are in homopolymeric regions (false positives), than other resistant variants.

Hypothesis 2B.1: A larger proportion of the population will have evidence of RT inhibitor associated resistance mutations using deep sequencing compared to standard bulk sequence analysis.

Hypothesis 2B.2: Deep sequencing and bulk sequencing results will agree for drug resistance mutations that occur frequently on Primer ID consensus sequences within a patient sample.

Hypothesis 2C.1: Results obtained using both ultra deep sequencing platforms will be highly concordant for mutations that occur on the majority of Primer ID consensus sequences.

Hypothesis 2C.2: Results obtained using both platforms will be discordant for mutations that lie within stretches of homopolymeric sequence, and for mutations that occur in very low abundance.

1.4 Specific Aim 3

Among study participants with HIV-1 RT sequences obtained using both bulk sequencing and the 454 deep sequencing platform, estimate the association of pre-therapy resistance to one or more antiretroviral agents within their first regimen on time-to-first virologic failure (VF).

Hypothesis 3.1: Participants harboring pre-therapy mutations conferring resistance to at least one ARV agent in their first HAART regimen will experience a shorter time-to-virologic failure compared to patients without any evidence of resistance at baseline.

Hypothesis 3.2: The magnitude of the association of ARV with resistance will be proportional to abundance, copies of the resistant variant or both.

CHAPTER 2 BACKGROUND

2.1 EPIDEMIOLOGY OF HIV INFECTION

The HIV epidemic has vastly altered the global infectious disease landscape since five patients were first identified with *Pneumocystis carinii* pneumonia in Los Angeles in October 1980 [27]. Over more than 30 years since, an estimated 34 million persons are living with HIV/AIDS worldwide, about 0.8% of the world's reproductive aged population [28]. Some regions of the world are disproportionately affected, such as sub-Saharan Africa, which accounts for 64% of the infected population [28]; however, even in the United States, over 600,000 people have died from AIDS and over 1 million people are living with HIV/AIDS [29]

Persons living with HIV in the US represent a paradigm shift in the course of infection. This shift occurred as HIV underwent a transition from an illness of long clinical latency leading inevitably to AIDS and death into a manageable chronic illness of unknown duration. This shift, which is unequally distributed within the US population of HIV-infected individuals, was primarily due to the advent of combinations of powerful antiretroviral drugs with different mechanisms of action against the HIV virus, or highly active antiretroviral therapy (HAART) [1, 2, 30]. More people with HIV/AIDS are living longer as increasingly potent and tolerable regimens have become available (Figure 2.1), but patients are completely dependent on HAART for remainder of their lives to prevent the onset of AIDS [31].

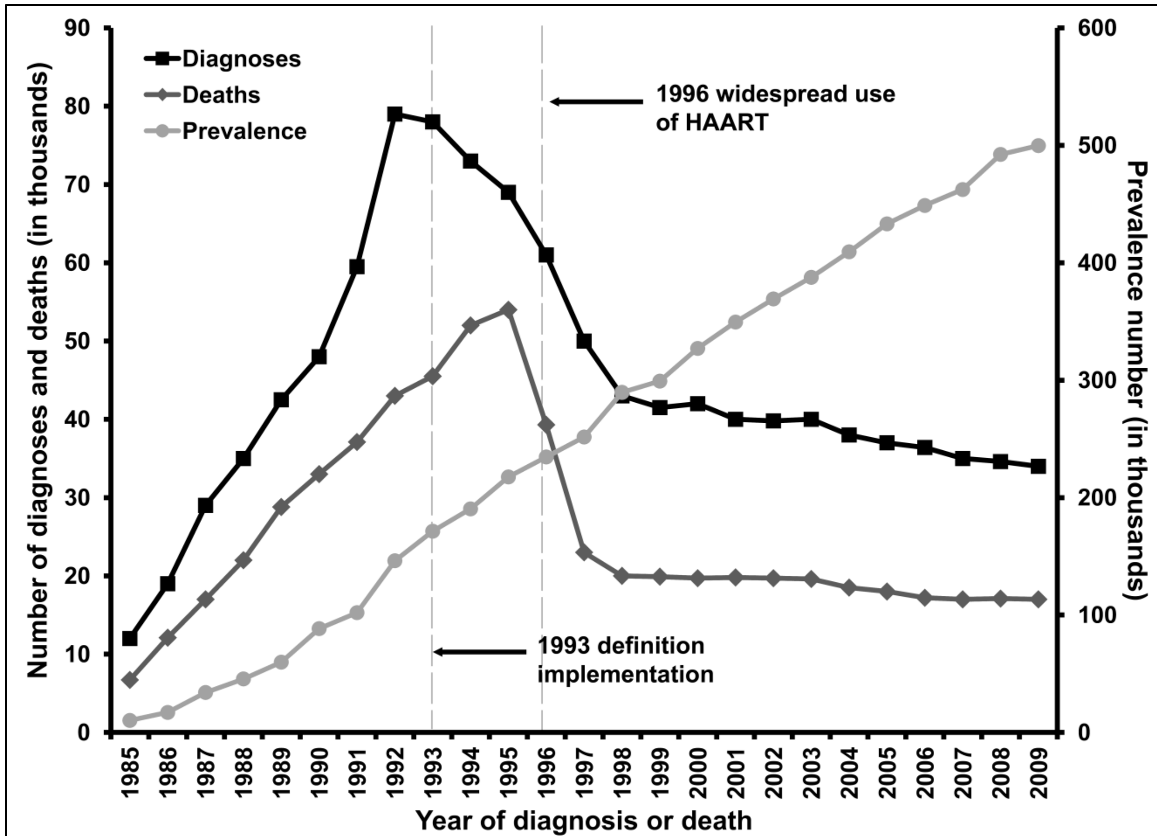

Figure 2.1. AIDS cases, deaths, and persons living with AIDS, 1985-2009, US [31]. Data is adjusted for reporting delays. In 1993, the Centers for Disease Control and Prevention revised the definition of AIDS to include persons with CD4 cell counts <200 cells/µL [32]. In 1996, widespread use of HAART began. HAART=highly active antiretroviral therapy.

After HIV was eliminated from the blood supply, HIV/AIDS was thought to be confined to high risk groups, particularly men who have sex with men (MSM) and injection drug users (IDU) [33]. Since peaking at 31% in 1993, the proportion of AIDS diagnoses attributable to IDU declined to 13% in 2011, while MSM accounted for about half of diagnoses during that same period [29]. AIDS cases among persons reporting high risk heterosexual contact, while decreasing since 2008, have increased proportionately overall from 11% in 1993 to 31% in 2011 [29]. The most striking shift in the US epidemic, however, is the widening disparity between whites and racial and ethnic minorities. Minorities accounted for fewer than half of all new AIDS diagnoses in 1985, but this proportion climbed to 74% in 2011 [29].

The Deep South, which includes Louisiana, Mississippi, Alabama, Georgia, South Carolina, and North Carolina [34], is a high priority region for HIV research, since the burden of AIDS is disproportionately high in that locale and changes in racial disparity and mode of transmission in the Deep South are at least as extreme as in any other region in the US. Nearly 17% of persons living with AIDS reside in these 5 states, which only account for 11% of the US population [29]. The Deep South has experienced a slower decline in AIDS cases since the introduction of HAART compared to other southern states and the remainder of the US: the Deep South experienced about a 19% decrease in AIDS diagnoses since 1996, while the remaining southern and non-southern states saw 38% and 52% declines [29, 35-45]. African Americans made up 70-77% of new AIDS cases in the Deep South in 2000-2004, which was higher than any other US region [46]. Further, the HIV epidemic in this region is driven by sexual transmission with heterosexual transmission playing an important role, particularly among women. Through 2002, 53% of all AIDS cases in the region were attributed to heterosexual transmission, with the majority of which were women [47].

2.2 HIV Virology

Acquired Immunodeficiency Syndrome (AIDS) was first recognized in 1980, but the causative pathogen, human immunodeficiency virus (HIV), was not identified until 1983, when scientists at the Pasteur Institute in Paris isolated a retrovirus from a lymph node of a patient with AIDS [48]. Since this discovery, HIV virology has been vigorously studied in hopes of developing effective vaccines, powerful treatments, and a potential cure.

Human immunodeficiency virus (HIV) belongs to the genus *Lentivirus* of the family Retroviridae based on shared morphology and similar mode of replication [49]. HIV is further classified into HIV-1 and HIV-2 based on genetic variability. HIV-1, which

originated in eastern and central Africa, and HIV-2 from western Africa, were both important in the early epidemic long before the 1980's [50], but the current pandemic is due to HIV-1, and in particular HIV-1 Group M [50, 51]. While HIV-1 Group M subtype C is the most prevalent in the world given its dominance in Africa, subtype B, which is most prevalent in North America and Western Europe, has traditionally been the most frequently studied.

HIV-1 is an enveloped virus with a matrix protein that surrounds a nucleoid structure in the center of the viral particle [49]. The nucleoid structure of mature HIV-1 contains the viral dimeric RNA genome surrounded by a nucleocapsid core, which in turn is housed in the viral capsid. This nucleoid structure also contains viral enzymes integrase, reverse transcriptase (RT), and protease (PR). Surrounding the virion is a host-derived envelope, which is embedded with numerous spike-like glycoproteins that are responsible for host cell attachment and fusion [52]. Each glycoprotein is composed of three transmembrane gp41 subunits (trimer) plus a gp120 trimer that binds to the host cell's CD4 receptor and CCR5 or CXCR4 coreceptor [53].

The HIV genome is approximately 9 kb in length and encodes ten genes: regulatory genes *tat* and *rev*; accessory genes *vpu*, *vpr*, *vif*, and *nef*; and structural genes *gag*, *pro*, *pol*, *env* (Figure 2.2). In addition, the HIV DNA genome includes non-coding regulatory sequences, including two flanking long terminal repeats (LTR). Among the coding regions, regulatory gene *rev* encodes a protein that shuttles incompletely spliced RNA transcripts from cellular nucleus to cytoplasm for translation and packaging, while *tat* encodes a transcription factor that upregulates HIV DNA expression [54]. Accessory genes *vpu*, *vpr*, *vif*, and *nef*, encode proteins that likely inhibit cellular retroviral defenses [49]. G*ag* and *env* encode proteins that make up the nucleoid structure and envelope, respectively, while *pro* encodes HIV-1 PR. Finally, *pol* encodes

enzymes necessary for reverse transcription of viral RNA into DNA (RT and ribonuclease H, RNAse H) and integration (integrase, IN) of viral DNA into the host genome. In this study, HIV-1 *pol* is the region of interest since the most widespread therapeutic strategies target its products.

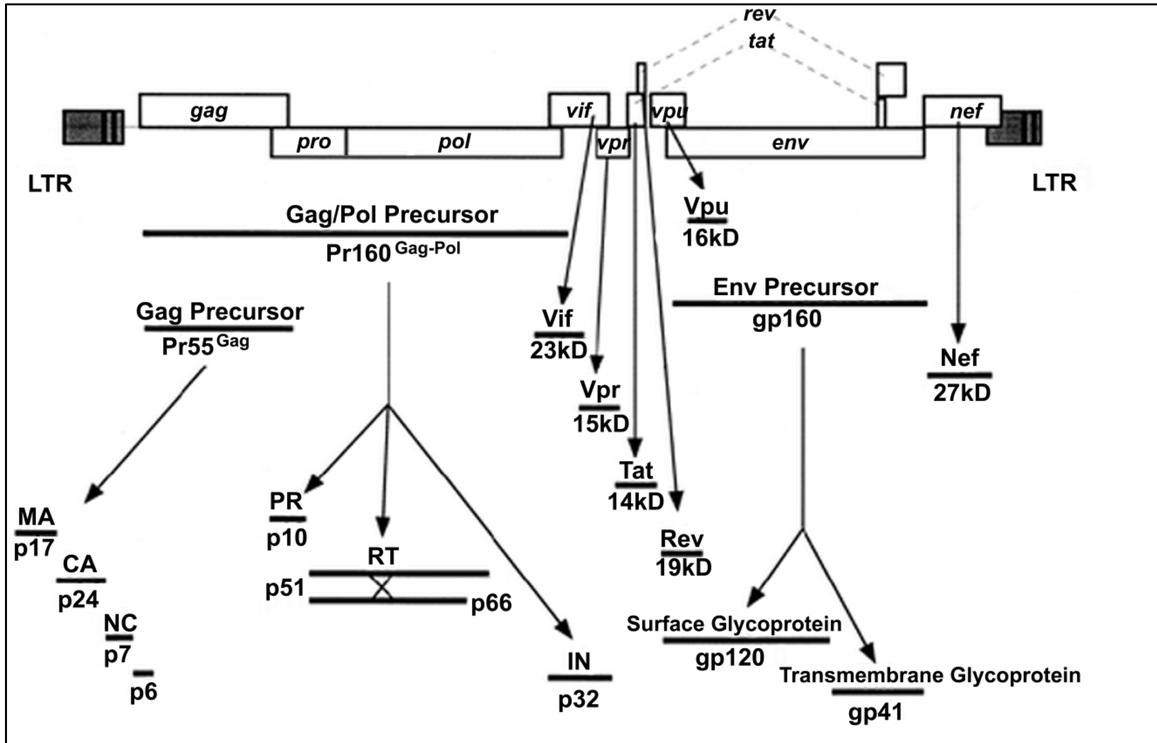

Figure 2.2. HIV-1-encoded proteins (reprinted from [49]). Location of HIV-1 genes, sizes of primary translation products, and processed mature viral proteins are indicated. LTR=long terminal repeat; MA=matrix protein; CA=capsid protein; NC=nucleocapsid protein; IN=integrase.

2.3 NATURAL HISTORY OF HIV INFECTION

The classic model for the course of HIV infection in an *untreated individual* can be divided into three distinct phases: acute infection; chronic infection; and AIDS (Figure 2.3) [55-57]. Typically, initial HIV infection is followed by a steep rise in plasma viral load and a drop in peripheral CD4+ T lymphocytes. This acute stage is followed by a partial rebound of CD4 cell count and a decline of plasma viral load to a relatively stable set-point. This period of stable HIV-1 RNA concentration is generally marked by the near absence of overt clinical illness; however, dynamic immunologic changes are taking

place within the infected individual. Eventually, CD4 cell count falls below a threshold, plasma HIV-1 RNA concentration slowly rises, and the individual begins to experience opportunistic infections. Immune function continues to deteriorate over time and death follows, usually within 2 years of the onset of opportunistic infections.

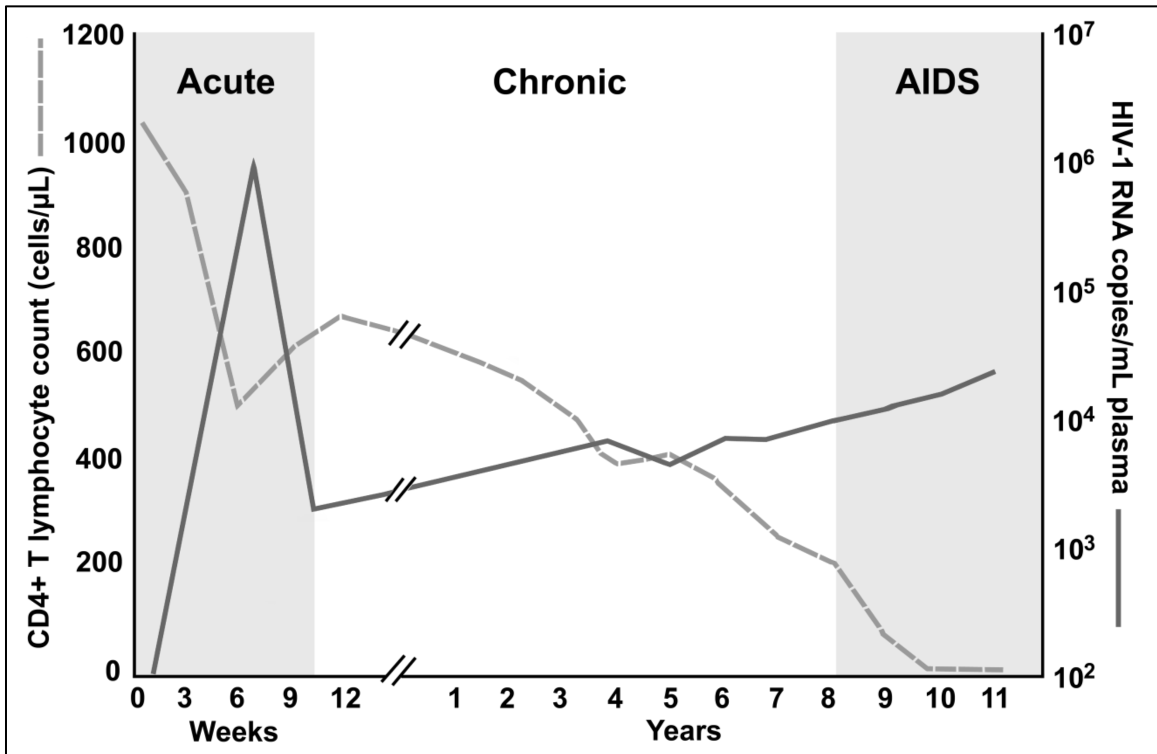

Figure 2.3. Classic model of HIV disease progression in the untreated HIV-infected adult. (adapted from [55-57]).

2.3.1 TRANSMISSION AND ACUTE/EARLY INFECTION

HIV disease begins with the transmission of the virus to a susceptible individual (time 0, Figure 2.3). HIV may be transmitted sexually, parenterally though exposure to infected blood products, or from mother to child. The probability of HIV transmission per exposure is highly variable and depends on exposure route, infectiousness of the transmitter, and susceptibility of the uninfected individual. Infectiousness of the transmitter is influenced by factors that alter the dose of virus to the susceptible individual. For example, acutely-infected individuals have higher viral loads in plasma, semen, and cervicovaginal fluids compared to persons with chronic HIV infection and

are therefore more likely to transmit HIV [58-60]. Conversely, infectiousness decreases with suppressive HAART, which lowers viral replication in these same fluids [61], and its use among discordant couples reduces HIV transmission [62, 63].

Susceptibility may be influenced by factors that disrupt the mucosal barrier of the susceptible individual (concurrent infection with other sexually transmitted diseases (STD)), that are related to or modify the number of susceptible cells at the site of infection, (male circumcision, CCR5 Δ32 allele), or that modulate immune response (human leukocyte antigen class I alleles) [64-67]. Individual effects of each of these co-factors on the probability of transmission are difficult to distinguish, given uncertain timing of events in the transmission process and high correlation between HIV, other STD, and high-risk behaviors [67]. How these co-factors might affect the transmission and acquisition of different variants of HIV-1, including those harboring drug resistance mutations, is even less certain.

Once HIV is successfully transmitted, the acute stage of infection begins. During the first 2 weeks of acute infection, the virus disseminates from local to peripheral lymphatic tissues, creating a reservoir from which the virus can persist and replicate for the duration of disease [68]. Establishment of this viral reservoir within the peripheral lymphatic tissues is accompanied by irreversible and massive depletion of effector memory CCR5+ CD4+ T cells, particularly within the gastrointestinal tract [69, 70]. Massive die off of CCR5+ CD4+ T cells is not reflected in peripheral CD4 cell counts, since CCR5+ CD4+ T cells are not as common in peripheral blood and lymphatic tissues where central memory CCR5- CD4+ T cells dominate [71].

Within infected CCR5+ CD4+ T cells, HIV replication increases exponentially without significant control by an adaptive immune response, resulting in a burst in plasma virus concentration often higher than 10 million copies/mL [72]. As the adaptive

immune system becomes activated, many individuals with acute HIV may experience a constellation of non-specific, flu-like symptoms, collectively called acute retroviral syndrome [73, 74]. The onset of acute retroviral syndrome and the steep rise in viral titers coincides with the initiation of a CD8+ cytotoxic T lymphocyte response that is ultimately insufficient to eliminate HIV infection [75], but often allows reduction of viral load to a lower, somewhat more stable set-point [76]. Both lower viral load set-point and more rapid viral decay after reaching peak viral load during acute HIV infection have been linked to increased duration of the chronic stage of infection [76].

2.3.2 Chronic Infection and AIDS

The chronic stage of HIV disease begins with the establishment of the viral load set-point and detection of antibodies by standard clinical assays (seroconversion), but the duration of this relatively asymptomatic stage differs substantially among individuals. Some fast progressors experience AIDS onset within 2-3 years of seroconversion, while some long term non-progressors maintain high CD4 cell counts, lower viral loads, and remain asymptomatic for greater than 15 years after seroconversion [77]. However, on average, the time between seroconversion and the onset of AIDS-defining clinical conditions is around 10 years [78]. Although chronic infection of any duration is a period of clinical latency, in the absence of therapy, immunologic changes occur within the infected host that ultimately lead to the collapse of the immune system and the onset of AIDS.

Once an infected person's CD4 cell count has declined below 200 cells/μL, the individual meets the CDC laboratory definition of AIDS onset [32]. At this level of immunodeficiency, there is profound damage to the architecture of the lymphatic structures [55, 79], probably resulting from both indirect damage caused by general chronic immune activation and direct damage due to cytopathic effects of HIV

replication and immunologic responses specific to HIV [56]. By the time HIV-infected individuals have met this definition of AIDS onset, many will already have begun experiencing symptoms such as oral candidiasis, herpes zoster outbreaks, recurrent oral ulcers, or constitutional symptoms such as prolonged fever, weight loss, or chronic diarrhea [32]. However, usually within 2 years after the laboratory threshold is reached, most begin experiencing the opportunistic illnesses that are listed in CDC category C, known as AIDS-defining clinical conditions [80, 81]. The number, duration, and severity of opportunistic infections vary among individuals, as does the prognosis. In one study, median survival for systemic illnesses such as progressive multifocal leukoencephalopathy was only 2 months (inter-quartile range, IQR: 1-5), while median survival for more superficial infections such as extrapulmonary tuberculosis was 19 months (IQR: 7-37) [82].

2.4 TREATMENT STRATEGIES FOR HIV INFECTION

Potent combinations of ARV agents (HAART) have lengthened the chronic stage of HIV infection, reduced mortality [1, 2, 30, 83, 84], and improved AIDS survival time [85]. The expanding ARV repertoire in clinical use now includes 26 drugs from five classes distinguished by their mechanism of action against HIV (Figure 2.4) [86]. The five classes are: nucleoside and nucleotide reverse transcriptase inhibitors (NRTI); non-NRTI (NNRTI); protease inhibitors (PI); entry inhibitors (EI), including fusion and CCR5 inhibitors; and integrase strand transfer inhibitors (InSTI). Development of newer, better ARV is guided by our increasing understanding of the life cycle of HIV-1 and fueled by our need to outpace the development of resistance and cross-resistance within classes among individuals taking these drugs.

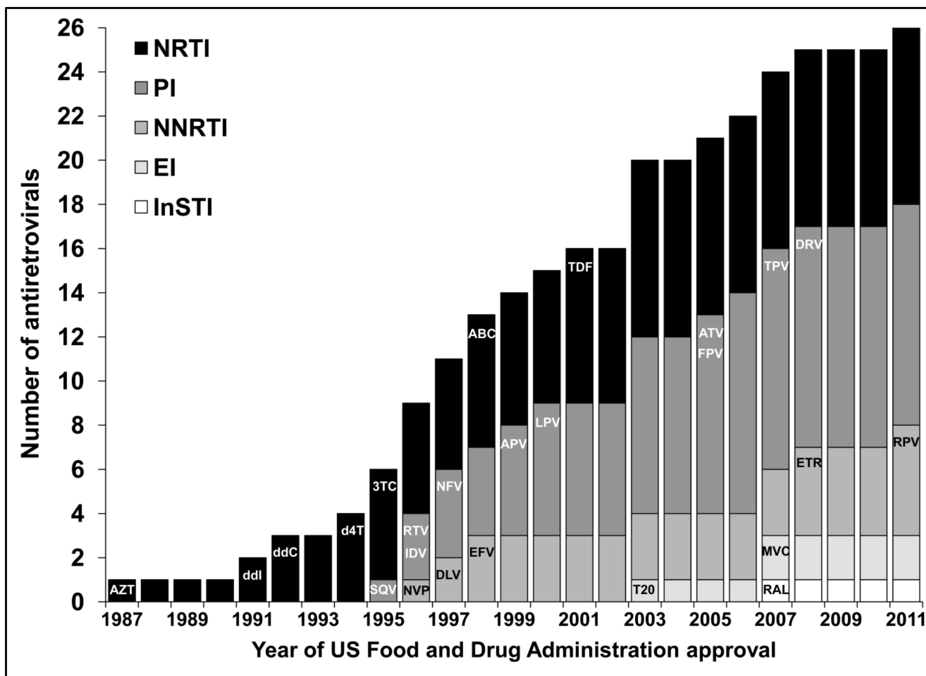

Figure 2.4. Antiretrovirals approved by the US Food and Drug Administration for the treatment of HIV-1 infection, 1987-2011 [86]. NRTI=nucleoside reverse transcriptase inhibitor; PI=protease inhibitor; NNRTI=non-NRTI; EI=entry inhibitor; InSTI=integrase strand transfer inhibitor; AZT=zidovudine; ddI=didanosine; ddC=zalcitabine; d4T=stavudine=3TC=lamivudine; ABC=abacavir; TDF=tenofovir; SQV=saquinavir; RTV=ritonavir; IDV=indinavir; NFV=nelfinavir=APV=amprenavir; LPV=lopinavir; ATV=atazanavir; FPV=fosamprenavir; TPV=tipranavir; DRV=darunavir; NVP=nevirapine; DLV=delavirdine; EFV=efavirenz; ETR=etravirine; RPV=rilpivirine; T20=enfuvirtide; MVC=maraviroc; RAL=raltegravir.

The primary goals of HAART, as listed in DHHS treatment guidelines, are to: (1) reduce HIV-related morbidity and mortality; (2) improve quality of life; (3) restore and preserve immune function; (4) maximally and durably suppress viral load to below the limits of detection; and (5) prevent HIV transmission [6]. Modern HAART regimens include two NRTI as a “backbone,” along with a single NNRTI, a PI boosted with a low dose of the PI ritonavir (PI/r), an additional NRTI, or an InSTI [6, 87]; however, individuals may also maintain virologic suppression with a combination of an NNRTI and PI/r without a nucleoside backbone [88]. HAART is an effective and durable inhibitor of HIV replication since the combination of drugs from different ARV classes with different mechanisms of action minimizes the possibility of resistance to all three agents that make up the HAART regimen.

2.4.1 INHIBITORS OF HIV-1 REVERSE TRANSCRIPTASE

Since the focus of this study is on reverse transcriptase inhibitor (RTI) resistance mutations, this discussion will focus on drugs within the NRTI and NNRTI classes. NRTI (Figure 2.4, black bars) are chain terminators of reverse transcription that halt reverse transcription of viral RNA when incorporated into the nascent DNA strand [89]. NRTI are the least potent inhibitors of viral replication because they are competitive inhibitors of naturally occurring cellular nucleotides and must be processed by the cell into an active form. Nucleo*side* reverse transcriptase inhibitors, including zidovudine (AZT), didanosine (ddI), discontinued zalcitabine (ddC), stavudine (d4T), lamivudine (3TC), abacavir (ABC), and emtricitabine (FTC) [86], lack all three phosphate groups that are needed for incorporation into a DNA molecule and must have 3 phosphates added by cellular kinases. In contrast, the nucleo*tide* reverse transcriptase inhibitor, tenofovir (TDF) [86], already has 1 phosphonate group prior to cell entry and require only 2 intracellular phosphorylation steps [90]. Once converted into an active form, NRTI halt DNA synthesis since they lack a 3’ hydroxyl group needed to connect the next nucleotide.

NNRTI (Figure 2.4, medium gray bars) are a more structurally diverse group of compounds that inhibit HIV-1 RT by binding a hydrophobic pocket near the active site of this enzyme [91]; thus, NNRTI are allosteric inhibitors of HIV-1 RT. Once bound to RT, NNRTI induce a conformational change at the active site of RT that prevents the enzyme from binding correctly to nucleotides, the RT primer, and the RNA template [91]. NNRTI are highly specific to HIV-1 RT, which allows the inhibitory concentration of NNRTI to be lower than the amount that would be toxic to the individual [91]. NNRTI have higher oral bioavailability compared to nucleoside analogs since they do not have to be converted by the body into an active form, and these compounds have longer serum half-lives than nucleoside analogues (NRTI). Currently-approved NNRTI include delavirdine (DLV),

nevirapine (NVP), efavirenz (EFV), etravirine (ETV), and most recently, rilpivirine (RPV) [86].

2.4.2 Timing of Initial Antiretroviral Therapy

When to initiate HAART and which ARV to combine is a complicated decision that is made based on individual and clinician considerations, including the presence of pre-existing resistance mutations [87]. Deferring HAART may benefit patients by: reducing the risk of treatment related toxicities and side effects; preserving and increasing future treatment options; decreasing the risk of treatment fatigue; and allowing patients to better understand the demands of HAART [6, 87]. On the other hand, deferring HAART may harm patients and their susceptible partners by: allowing irreversible damage to the immune system; increasing the risk of disease progression and death; and increasing the risk of HIV transmission to uninfected partners [6, 63, 87]. Given the risks and benefits of initiating HAART, current guidelines generally endorse initiation of therapy in all individuals with HIV infection, irrespective of disease stage [6]. All treatment guidelines recommend therapy initiation when individuals become symptomatic, regardless of CD4 cell count or HIV-1 RNA level. In asymptomatic patients, however, CD4 cell count is favored over HIV-1 RNA level in determining antiretroviral therapy initiation. Individuals with CD4 cell counts 200 cells/µL or below are offered therapy since initiation below this threshold increases the risk of disease progression or death and reduces the efficacy of treatment [92]. Therapy initiation at CD4 cell counts between 200 and 350 cells/µL is also recommended, since deferring therapy until CD4 cell counts dip below 200 cells/µL is associated with increased mortality and incidence of opportunistic infections [6, 87, 93]. Finally, for patients with CD4 cell counts ranging between 351 and 500 cells/µL, initiation of therapy is also recommended since there is good evidence that disease progression is delayed among these individuals, even if mortality is not clearly reduced [6]. CD4 cell count driven

recommendations underscore the importance of earlier diagnosis, which unfortunately is not the norm in the Southeastern US setting [94].

2.4.3 Virologic Response to Initial Antiretroviral Therapy

Once HIV-infected individuals begin HAART, plasma HIV-1 RNA decays in at least biphasic fashion. Patients first experience a 2 to 3 $\log_{10}$ decrease in HIV-1 RNA plasma concentration within 2 to 4 weeks of starting therapy [95, 96]. This first phase of rapid decline is followed by a slower, second phase of decay in which plasma HIV-1 RNA levels fall below detection limits within an additional 2 to 24 weeks of suppressive therapy [97-101]. It is now clear that this second phase lasts indefinitely, since even in the presence of potent HAART, there is evidence of ongoing viremia below the limits of detection in clinical HIV-1 RNA quantification assays [98, 102, 103]. Also, many patients who achieve HIV-1 RNA levels below detection limits experience transient episodes of detectible HIV-1 RNA, or “blips.” Each viral blip is probably associated with its own biphasic decline [104], but since blips are usually detected in the context of clinical care, where HIV-1 RNA level is measured every 4 to 8 weeks, only a single point on the curve is captured. Intermittent viremia may have important implications for drug resistance, since ongoing replication of virus under antiretroviral pressure may select for resistant variants that may be archived in resting cells or compartmentalized in tissues where viral dissemination is restricted, allowing for persistence of drug resistant variants within the treated individual [105, 106]. Episodes of intermittent viremia after initial suppression occur in 20-40% of patients in clinical care within the first year of therapy [107-109], and most of these patients return to suppressed levels upon their next viral load measurement. However, a small proportion of patients with intermittent viremia (7-15%) do not return to previously suppressed HIV-1 RNA levels by their next HIV-1 RNA measurement and experience virologic failure, a form of treatment failure [107-109].

2.4.4 TREATMENT FAILURE

Treatment failure occurs when patients on therapy do not experience expected improvements in terms of virologic response, immune reconstitution, and/or HIV disease progression [6]. Treatment failure may be the result of poor adherence, insufficient drug concentration, advanced AIDS or other comorbid conditions, or drug resistance. Failure often leads to modification or discontinuation of ART, which is common in the clinical setting: an estimated 48% of patients in routine clinical care in the UK switched one ARV in their regimen within two years of HAART initiation [110]. In a Southeastern US cohort, 55% of patients discontinued their first HAART regimen within one year; however, most of these were associated with drug toxicity rather than treatment failure [111, 112].

For this study, one type of treatment failure will be defined – virologic failure (VF). Although virologic and immunologic response are highly correlated, there are many cases in which patients achieving virologic suppression fail to achieve immunologic benefits from HAART [113]. Also, there are many cases in clinical care where patients may experience immunological benefits despite persistent viremia while on HAART [114]. In most settings, patients experiencing VF include individuals who either fail to achieve a significant decline in viral load within the first 6 months of HAART (0.5 to 1 $\log_{10}$ copies/mL) or have detectible HIV-1 RNA levels after 6 months of HAART. HIV-1 RNA assay detection limits have evolved over time, and this cut-off is usually set for the highest detection limit of all assays used. This definition of failure is flexible since it allows for differential viral decay after initiation of HAART that occurs between individuals [95]; it includes those individuals who never achieve suppression as well as those experiencing virologic rebound; and it allows for viral load blips, which may not be clinically important [108, 115].

2.5 Selection of Resistance Mutations with Antiretroviral Therapy

Treatment guidelines emphasize the importance of viral suppression to below the limits of detection primarily to prevent evolution of antiretroviral resistance mutations [6]. With selective pressure of non-suppressive therapy, resistant viral variants may acquire additional mutations that increase resistance to a single drug or lead to cross resistance to other drugs within that class. Alternatively, resistant variants may acquire mutations that compensate for penalties to viral fitness incurred with initial drug resistance mutations [116]. Variants with resistance mutations may also recombine, producing populations with additional resistance mutations on the same genome [117, 118].

Following convention, resistance mutations are indicated using single letter amino acid abbreviations and the codon number; for example, M184V indicates a change from wild type methionine (M) to resistant valine (V) at codon 184 of HIV-1 RT (Appendix 2.1). Most known amino acid changes and insertions or deletions associated with NRTI or NNRTI resistance fall within codons 41 and 225 of HIV-1 RT [119]. The accumulation of resistance mutations within these regions is no accident: changes in these codons often confer structural changes that disrupt substrate binding and processing, not only for natural substrates, but also for inhibitors.

2.5.1 Antiretroviral Resistance and Viral Fitness

Viral fitness is defined as the virus's capacity to replicate in a given environment [120]. Since fitness is contextual, the microenvironment in which HIV replicates must be well defined. Further, the concept of fitness is so broad that it may be altered at virtually any stage in the viral life cycle. Multiple genes may be in play, thus complicating comparisons of viral fitness across studies that use different methods to measure viral fitness. However, the concept of viral fitness is central to determining the relative abundance of different viral variants in the individual; it is expected that in a given microenvironment, such as the presence of non-suppressive therapy, variants with a

fitness advantage allowing them to replicate more efficiently will emerge as the dominant population.

In the absence of ARV, many viruses with resistance mutations are less fit than the wild type virus, since many of the mutations introduce conformational changes to the target enzymes that negatively impact function. The fitness difference between resistant and wild type variants differs substantially by ARV class, location, type and number of resistance mutations, and the methods and environment in which fitness is measured [121]. Because of the fitness cost associated with many drug resistance mutations, the more fit, drug sensitive virus tends to re-emerge as the dominant population once drug pressure is removed, assuming the patient was initially infected with sensitive virus [9]. Such re-emergence includes replacement of resistant variants with archived sensitive variants, but reversion of resistance through back mutation is also possible.

NRTI mutations accumulate under suboptimal therapy, primarily increasing resistance rather than compensating for a loss in fitness. Many NRTI mutations (A62V, K65R, T69ins, V75I, L74V, F77L, Y115F, F116Y, and Q151M) accumulate near primer binding site of RT, impairing the function of the enzyme. For example, the K65R amino acid change is associated with decreased incorporation of natural nucleoside substrates by HIV-1 RT, especially dATP [122]. RT function may also be impaired by a mutation within the polymerase active site, M184V/I, which increases fidelity and reduces processivity of the enzyme [123]. A separate class of mutations, selected by AZT and d4T, occur outside the primer binding site in “palm” and “finger” subdomains of RT, and may have less of an effect on fitness than mutations occurring near the polymerase active site [124]. These are collectively called thymidine analogue mutations (TAM), and include M41L, D67N, K70R, L210W, T215Y/F and K219Q/E. During non-suppressive AZT or d4T therapy, TAM tend to accumulate in two exclusive patterns, with an

increasing number of mutations associated with increased resistance and eventually cross-resistance to other nucleoside or nucleotide analogues within the NRTI class [125, 126]: (1) M41L/L210W/T215Y (TAM-1); or (2) D67N/K70R/T215F/K219Q (TAM-2). Clustering of TAM into these two patterns has been attributed to changes in fitness determined largely by T215Y and L210W [126].

Unlike many NRTI resistance mutations, amino acid changes conferring resistance to NNRTI are not often associated with large fitness reduction in the absence of drug. NNRTI mutations, including K103N, Y181C, and Y188L, are selected rapidly under failing NNRTI-based therapy. They may be associated with high-level, class-wide resistance, and they may persist for lengthy amounts of time once drug pressure is removed [127, 128]. NNRTI mutations cluster near the hydrophobic pocket of RT that is the target for NNRTI, and most do not interfere with RT function. However, some NNRTI associated resistance mutations, including V106A, G190S/E, and P236L, *are* associated with a significant decline in replicative capacity in the absence of NNRTI; these effects on fitness are believed to be due to a decline in RNase H activity of mutant RT [129].

2.6 Mechanisms of Reverse Transcriptase Inhibitor Resistance

Pathways of ARV resistance vary by drug class, by ARV within the class, and by individual mutation. Many of the best characterized mechanisms were first described under monotherapy, but mutations may accumulate and interact in more complex ways under combinations of ARV from multiple classes (HAART) [130]. Further, many resistance mutations to one class may interact in unexpected ways with mutations to another class, complicating expectations about the evolution of resistance mutations under suboptimal therapy.

NRTI resistance mutations were reported almost immediately after the earliest NRTI became available [131]. However, as other analogues were introduced, the

spectrum of NRTI mutations also broadened, revealing two mechanisms of NRTI resistance: (1) reduced incorporation of the NRTI and (2) ATP-mediated excision of the NRTI from the terminated DNA strand [116]. The first mechanism, whereby RT shows a reduced capacity to add NRTI to the DNA chain, is associated with M184V/I, K65R, and the Q151M complex. M184V/I decreases incorporation of 3TC/FTC since the side chains of I/V sterically hinder binding of 3TC or FTC [132]. K65R, associated with ABC, ddI, and TDF resistance, occurs within the fingers domain of RT, where the wild type K residue forms a bridge with an incoming nucleotide; this bridge is disrupted by replacement of K with R [133, 134]. Q151M, which is associated with class-wide resistance upon accumulation of four additional mutations [135], appears to decrease binding of nucleo*side* analogues [136].

The second mechanism of resistance to NRTI involves excision of the NRTI from the 3' end of the DNA strand, or primer unblocking. This mechanism is associated with TAM, and involves the use of cellular ATP by mutant RT to attack the bond between the NRTI and the terminated DNA strand [116]. This second mechanism may be associated with cross-resistance to drugs within the NRTI class [137]. Excision of NRTI may also be less efficient in the presence of M184V, which may antagonize primer unblocking because of its position at the active site [138].

Pathways for NNRTI resistance mainly involve mutations that occur in the hydrophobic pocket of RT. This hydrophobic pocket, which is distal to the polymerase active site, is only formed once the NNRTI is bound to RT. In wild type unbound RT or RT bound with DNA, the side chains of tyrosine residues at codons 181 and 188 fill the pocket [139, 140]. Once the inhibitor binds, 181 and 188 side chains rotate toward the polymerase active site, and the hydrophobic pocket is formed. Mutations in this region confer resistance by: (1) blocking entry of the NNRTI; (2) inhibiting interactions between

the NNRTI and the amino acids within the pocket; and/or (3) changing the conformation of the pocket [141]. For example, the K103N substitution may prevent binding of NNRTI since the mutant N residue allows additional hydrogen bonds that prevent the hydrophobic pocket from opening [142]. Y181C, however, increases the size of the hydrophobic pocket of the enzyme, preventing first generation NNRTI, such as NVP and DLV, from making critical contacts with residues inside the pocket [143]. Newer NNRTI, such as RPV and ETR, are less affected by this mutation since they are more flexible and less dependent on interaction with 181 and 188.

2.7 Epidemiology of Acquired Antiretroviral Resistance

Accumulation of resistance mutations on therapy is a gradual process, leading to cross resistance and eventually exhaustion of drugs within a class, particularly as patients fail multiple HAART regimens [144]. Drug resistance that is acquired on therapy accounts for the majority of drug resistance in the HIV-infected population. The patterns of resistance among treated patients are somewhat reflected among therapy-naïve patients [145]. Comparing resistance across studies is difficult since only a fraction of patients at risk are tested, mutations considered relevant change over time or vary across mutations lists, and estimates must be updated as treatment strategies evolve. Further, standard resistance assays do not detect minority drug resistant variants [146], and resistance tests obtained at a single time point may not reflect the full spectrum of resistance within the individual, since mutations may have receded to undetectable levels in the absence of drug selection pressure [147].

2.7.1 Prevalence of Acquired Reverse Transcriptase Inhibitor Resistance

Even with the above caveats, multiple efforts have been made to estimate the burden of resistance among patients failing therapy, particularly in regions where HIV treatment is widely available. Among a representative sample of viremic patients receiving care in the United States through 1998, the overall prevalence of resistance

was estimated to be 76%; 71 and 25% of patients showed reduced susceptibility to at least one NRTI or NNRTI, respectively [148]. These estimates are from the early HAART era, when most patients in care initiated therapy with mono- or dual-NRTI regimens (non-HAART). Since then, the pool of patients on therapy is increasingly dominated by those who initiate treatment with HAART, changing the pattern of resistance over time.

Temporal associations demonstrate that resistance trends coincide with changes in ARV usage [149-154]. For example, in British Columbia in 1996, nearly 90% of patients failing therapy harbored resistance to at least one ARV within one class [149]. By 2003, this number fell to 60%. However, while in 1996 only 14% of patients had evidence of dual class resistance, and virtually none had triple class resistance, these numbers rose to 39% and 7%, respectively, by 2003. These 2003 estimates of multi-class resistance were still dominated by patients who started with non-HAART regimens.

HAART initiators are expected to develop multi-class resistance less frequently than those initially treated with mono- or dual-NRTI therapy, since the former begin with a regimen with multiple mechanisms of inhibiting viral replication. For example, among viremic patients in North Carolina, 26% of patients who initiated therapy with a non-HAART regimen had evidence of triple-class resistance, while the prevalence was only 10% in those who started with HAART [155]. As more patients begin therapy with more potent and tolerable regimens, it is expected that complete class-wide resistance, in which patients have exhausted all options within a class, will remain rare (under 5%) [154, 156].

While the frequency of particular resistance mutations is strongly associated with ARV usage patterns, some unexpected patterns emerge, even if the agents known to select for rarer mutations are widely used. Among NRTI resistance mutations, the M184V amino acid substitution continues to be one of the most common mutations,

which is expected given pervasive use of 3TC and FTC in clinical care [149, 150, 155, 157, 158]. TAM are also common, especially T215Y/F and M41L, with 30-60% having evidence of either mutation [149, 150, 155, 157, 158]. K65R, in contrast, is usually seen in 5% or fewer patients failing therapy [149, 150, 153, 157-159]. K65R did appear to be on the rise after the introduction of TDF in 2001; yet, since 2005, its frequency has declined with decreased co-administration of TDF with ddI and ABC [160, 161]. The Q151M complex and the 69 insertion complex, associated with broad NRTI resistance, are also extremely rare [116]: in an Italian cohort of patients failing HAART, fewer than 3% and 1% had evidence of the Q151M and 69 insertion complex, respectively [156].

For NNRTI associated resistance, the most common mutations detected among patients failing therapy are K103N and Y181C, and their frequency depends on the dominant NNRTI in use as well as the specific co-administered NRTI. In settings where EFV is dominant, K103N is the most common [150, 157], while in populations in which NVP is the NNRTI of choice, Y181C is detected more frequently or as much as K103N [149], which is only selected by NVP combined with d4T or AZT [162]. As EFV became more widespread in Italy, for example, accounting for 6% of ARV used in 1999 and 18% in 2003, the prevalence of K103N jumped from 17% to 29% [153].

2.7.2 Incidence of Acquired Reverse Transcriptase Inhibitor Resistance

Studies in which viremic patients were maintained on a failing ART regimen with two consecutive genotypic antiretroviral resistance tests (GART) have allowed estimation of the incidence of new drug resistance mutations as well as the identification of factors associated with their evolution. In a retrospective sample of 106 patients treated with a median of 6 ARV, 75% of patients maintained on a failing HAART regimen for a median of 14 months acquired a new resistance mutation, including 44% of those taking an NRTI, and 29% of those taking an NNRTI [163]. Results from the University of

North Carolina agree with these estimates: among 98 viremic patients with comparable ARV exposure history, 60% acquired a new drug resistance mutation over a median of 9 months of stable HAART; 38% and 46% of patients receiving an NRTI or NNRTI, respectively, acquired at least one new mutation within that class [164]. New NRTI resistance was primarily the M184V substitution, detected among 42% of patients on a regimen containing 3TC or FTC. Among less heavily treated patients in a European cohort (median 5 ARV), 77% of patients maintained on failing HAART for a median of 11 months acquired a new mutation [165]. At the other extreme, in a prospective sample of 106 heavily treated patients (median 8 previous ARV), 44% had evidence of a new mutation at 12 months [166].

Different estimates of resistant mutation incidence among these studies could be due to differences in the amount of drug resistance patients had at baseline, since most studies reported increased incidence was associated with fewer mutations at baseline. The study reporting the lowest incidence, for example, included only patients with at ≥1 major resistance mutation at baseline, while all other studies included a mixture of patients with and without evidence of resistance at baseline [166]. The effect of baseline mutations may be due to the evolutionary constraints on the virus, since additional mutations are often associated with a fitness cost to the virus and since a limited amount of selection pressure may be exerted by a particular regimen [21]. This ceiling effect may be critical for patients starting their first HAART regimen, since therapy-naïve individuals are expected to have fewer resistance mutations at baseline.

Studies also identified the amount of time spent on the failing regimen [163], lower nadir CD4 cell count [165], increase in HIV-1 RNA level while on the stable regimen [164, 165], and average HIV-1 RNA level between 3-4 $\log_{10}$ copies/mL [164] as risk factors for acquiring new resistance. For the latter, an average HIV-1 RNA level

below 3 $\log_{10}$ copies/mL may be associated with sufficient suppression of replication to prevent the selection of mutations into the dominant population. In contrast, an average HIV-1 RNA level above 4 $\log_{10}$ copies/mL is probably a marker for poor adherence and thus insufficient ARV concentration to select for resistance mutations [164]. However, even among patients who achieve suppression well below 3 $\log_{10}$ copies/mL, there is evidence that drug resistance mutations that may not be detectible with standard GART still emerge in patients on HAART. Among a small number of therapy-naïve patients starting their first HAART regimen, 40% of patients developed minority drug resistant variants during the first phase of decay even though they eventually achieved viral loads below the limits of detection [167].

2.7.3 Clinical Impact of Acquired Antiretroviral Resistance

At least in the short term, the accumulation of resistance mutations may be associated with some level of immunologic benefit, although less so than observed under complete suppression [114, 158, 168]. In one observational study among women who experienced virologic failure, those with evidence of baseline NRTI resistance and NNRTI resistance at failure still achieved an average increase of 118 CD4 cells/µL (95% CI: 38-198) after one year of HAART [158]. However, those with baseline NRTI resistance alone showed some evidence of a *decrease* in mean CD4 cell count of 31 cells/µL (95% CI: -82-20) at one year, while those who acquired NNRTI resistance alone experienced virtually no change. Increased benefit observed for women with both NNRTI and NRTI resistance could be due to the interaction between these mutations that causes hypersusceptibility to NNRTI [169].

In terms of virologic response, patients enrolled in a US clinical cohort (N=572) with fewer resistance mutations at baseline were more likely to achieve HIV-1 RNA <400 copies/mL over follow-up: 28% of those with 0-2 mutations, 21% of those with 3-6

mutations, and 10% of those with ≥7 mutations achieved suppression at 6 months [170]. As the repertoire of available ARV expands, the number of mutations may become less important than the activity of ARV within the regimen. In two cohort studies of highly treated patients with baseline resistance, patients receiving regimens containing ≥3 active ARV [171, 172] were more likely to achieve HIV-1 RNA <50 copies/mL within one year [173, 174].

Finally, there is also evidence that resistance mutations acquired during early treatment are linked to an increase in mortality. Patients enrolled in a British Columbia cohort experienced an increased hazard of all cause mortality with the emergence of any major resistance compared to patients without emerging resistance (hazard ratio (HR): 1.8, 95% CI: 1.3-2.4) [175]. Patients with emerging NNRTI resistance experienced the highest mortality rates (HR: 3.0, 95% CI: 2.0-4.6). These results were later confirmed in a randomized clinical trial estimating the effect of three different HAART strategies (NNRTI, PI, or NNRTI+PI) on the emergence of drug resistance over 5 years of follow-up among therapy-naïve patients [176]. Patients who failed and acquired NNRTI resistance were 2.4 times as likely to die or develop AIDS compared to patients who maintained suppression (95% CI: 1.5-3.7) [176]. Interestingly, patients who failed without any evidence of resistance also experienced excess mortality and AIDS incidence compared to patients who were suppressed (HR: 1.8, 95%CI: 1.2-2.7), which could be explained by their lower adherence scores and less time on therapy, indicating subtherapeutic exposure to HAART.

2.8 Epidemiology of Transmitted Reverse Transcriptase Inhibitor Resistance

A major consequence of acquired drug resistance among patients with detectible viral load is the increased risk of transmission to susceptible individuals. Acquired resistance has been shown to be transmitted sexually [177], from mother to child [178],

and parenterally [179]. Drug resistance transmitted to an HIV negative individual may also be further spread to other susceptible individuals [180]. Transmitted drug resistance (TDR) represents not only a public health threat, but it is also detrimental to individuals starting HAART since it may compromise initial treatment and set the patient on an accelerated path to multiple HAART failures and exhaustion of future treatment options. Further, TDR has been shown in many studies to persist for long periods of time in the absence of therapy, meaning reversion to sensitive strains may be unlikely in those initially infected with TDR, particularly when the resistant virus must pass through a less fit intermediate phase [13-15, 105, 128]. Resistant variants may also be found in resting CD4 cells of therapy-naïve individuals, so even if the dominant population circulating in peripheral blood is sensitive, resistant variants may re-emerge once the patient is exposed to the relevant ARV [16].

The epidemiology of TDR is complicated because there is no single standardized list of mutations or system of interpretation defining TDR. In fact, TDR may be thought of as a special case of pre-therapy resistance, since some mutations correlated with acquired resistance may be natural polymorphisms that have no effect on phenotypic resistance [181]. Other mutations may actually be markers for more extensive resistance among treated patients [182]. Consequently, estimates of TDR based on patients with acquired resistance, such as the list of mutations maintained by the International Antiviral Society-USA (IAS-USA) [183] or the online resistance database maintained by Stanford University [171, 184] may not reflect the true extent of TDR. For example, T215 revertant mutations are not usually included in these algorithms since they do not affect phenotypic resistance in vitro; however, these variants rapidly evolve to AZT resistance under selective drug pressure and are associated with treatment failure. The presence of T215 revertants reflects either transmission of T215Y/F and then reversion to a non-wild

type amino acid in the absence of therapy, or transmission of the revertant [185, 186]. Also, some resistance algorithms include V118I, which among patients failing therapy occurs in tandem with TAM-1 mutations. However, among therapy-naïve individuals starting AZT or 3TC regimens, V118I alone has no effect on virologic or immunologic response compared to patients with sensitive virus [187].

To address issues of TDR surveillance, a list of mutations has been developed by Shafer and colleagues, and implemented by the World Health Organization (WHO) (Appendix 2.2) [188, 189]. The selection process for this list begins with lists of known resistance mutations from expert opinion and available clinical data. Polymorphisms on these lists present in >5% of sequences from therapy-naïve individuals are excluded to preclude the inclusion of naturally occurring mutations that do not necessarily reflect transmission events. The list of surveillance drug resistance mutations (SDRM) is then simplified to maximize sensitivity and specificity for probable TDR.

Even if resistance mutations detected among therapy-naïve individuals represent transmission events, TDR estimates are often plagued by selection bias, since resistance testing may be selectively applied to individuals perceived as more likely to harbor resistance, especially in resource poor settings [182]. Estimates may be influenced by the stage of infection at testing, since less fit resistant variants may wane over time to clinically-undetectable levels without ARV selective pressure [146]. However, even when resistant variants may still dominate, such as among those with recent HIV infection, approximations may be biased since these individuals often have symptomatic acute HIV infection, undergo frequent HIV testing, or are identified by their partners on therapy, thus threatening generalizability of TDR estimates [190].

2.8.1 PREVALENCE OF TRANSMITTED DRUG RESISTANCE

Given these challenges, some generalizations about TDR may still be drawn from available observational information. TDR, analogous to acquired resistance, is strongly correlated with ARV use in the treated population. TDR is more likely to be detected among whites, MSM, individuals with subtype B infections, and among those with partners on therapy [180, 191-194], reflecting a correlation with greater access to ARV and more frequent testing behaviors. TDR is most common where ARV use has been more widespread, while the prevalence remains low where antiretroviral coverage is lowest [182, 195]. TDR more common among individuals with recent HIV infection (Table 2.1) compared to individuals with chronic infection or infection of unknown duration (Table 2.2), probably reflecting both recession of less fit resistant variants over time and overrepresentation of groups more likely to carry TDR, such as white MSM whose sexual partners may be on non-suppressive therapy [195].

Table 2.1. Reverse transcriptase inhibitor resistance among antiretroviral-naïve individuals with recent HIV-1 infection.

Region	Years	N	Resistance Definition	Prevalence % NRTI	% NNRTI	Reference
Africa						
Côte d'Ivoire	2002-2006	100	IAS-USA	3.0%	2.0%	Toni et al.[196]
Kenya	2009-2010	68	SDRM[§]	1.5%	7.4%	Sigaloff et al.[197]
Uganda	2004-2010	72	SDRM[§]	0.0%	1.4%	Ssemwanga et al.[198]
Uganda	2007	70	SDRM[§]	2.9%	4.3%	Ndembi et al.[199]
Uganda	2009-2010	47	SDRM[§]	0.0%	6.4%	Nazziwa et al.[200]
Kenya	2006-2009	64	SDRM[§]	1.6%	1.6%	Price et al.[201]
Rwanda	2006-2009	78	SDRM[§]	0.0%	3.8%	Price et al.[201]
South Africa	2006-2009	5	SDRM[§]	0.0%	20.0%	Price et al.[201]
Uganda	2006-2009	92	SDRM[§]	1.1%	2.2%	Price et al.[201]
Zambia	2006-2009	169	SDRM[§]	1.8%	1.2%	Price et al.[201]
Asia						
China	2006	53	SDRM[§]	0.0%	1.9%	Zhang et al.[202]
Thailand	2003-2006	305	IAS-USA	2.0%	2.3%	Apisarnthanarak et al.[203]
Europe						
7 Countries	1987-2003	438	IAS-USA[†]	10.5%	3.4%	Masquelier et al.[204]
France	2001-2002	303	IAS-USA[†]	10.3%	3.3%	Descamps et al.[12]
Germany	1996-2007	1,276	SDRM[§]	7.5%	3.5%	Bartmeyer et al.[205]
Italy	2000-2010	226	SDRM[§]	7.0%	7.0%	Colafigli et al.[206]
Romania	1997-2011	5	SDRM[§]	0.0%	0.0%	Temereanca et al.[207]
Slovenia	2005-2010	31	SDRM[§]	6.5%	3.2%	Lunar et al. [208]
Spain	1997-2004	198	IAS-USA[†]	9.6%	4.0%	de Mendoza et al.[209]
Switzerland	1996-2005	822	SDRM[‡]	5.5%	1.9%	Yerly et al.[180]
United Kingdom	1996-2004	316	IAS-USA[†]	7.0%	4.1%	UK et al.[210]
United Kingdom	2004-2006	85	IAS-USA[†]	4.7%	2.4%	Booth et al.[211]
United Kingdom	2005-2006	140	IAS-USA[†]	0.7%	3.6%	Fox et al.[212]

Table 2.1 (continued). Reverse transcriptase inhibitor resistance among antiretroviral-naïve individuals with recent HIV-1 infection.

Region	Years	N	Resistance Definition	Prevalence % NRTI	% NNRTI	Reference
Latin America						
Argentina	2003-2005	28	IAS-USA	3.6%	3.6%	Dilernia et al.[213]
Brazil	1998-2002	55	IAS-USA†	9.1%	0.0%	Barreto et al.[214]
Brazil	1999-2001	51	IAS-USA	2.0%	0.0%	Varella et al.[215]
Cuba	2003-2011	194	SDRM§	4.6%	3.6%	Pérez et al.[216]
El Salvador	2008	19	SDRM§	0.0%	10.5%	Murillo et al.[217]
Mexico	2003-2005	10	IAS-USA	0.0%	0.0%	Viani et al.[218]
Peru	2002	33	IAS-USA	3.0%	3.0%	Lama et al.[219]
North America						
Canada	1996-2003	180	IAS-USA†	7.8%	3.9%	Jayaraman et al.[191]
Canada	2000-2001	221	IAS-USA†	5.9%	1.8%	Routy et al.[220]
United States						
7 Cities	1995-2006	1,311	SDRM‡	10.0%	7.0%	Liu et al.[221]
6 Cities	1999-2003	195	SDRM‡	8.7%	6.7%	Eshleman et al.[222]
15 Sites	2004	55	IAS-USA	3.6%	14.5%	Viani et al.[218]
Chapel Hill, NC	1999-2010	43	SDRM§	0.0%	18.6%	Yanik et al.[223]
New York	1995-2004	361	IAS-USA†	13.0%	8.0%	Shet et al.[224]
New York	1995-2010	600	IAS-USA	8.3%	6.8%	Castor et al.[225]

NRTI=nucleoside(tide) reverse transcriptase inhibitor; PI=protease inhibitor; NNRTI=non-NRTI; IAS-USA=International Antiviral Society–USA Panel; SDRM=surveillance drug resistance mutations. *Estimate includes intermediate- and high-level resistance only. †Modified to include 215 revertant mutations and exclude V118I mutation of HIV-1 reverse transcriptase. ‡Based on transmitted resistance surveillance mutation list assembled by Shafer et al. [188]. §Based on surveillance drug resistance mutations list updated in 2009 [189].

Table 2.2. Reverse transcriptase inhibitor resistance among antiretroviral-naïve individuals with HIV-1 infection of unknown duration.

Region	Years	N	Resistance Definition	Prevalence % NRTI	Prevalence % NNRTI	Reference
Africa						
6 Sub-Saharan Countries	2007-2009	2,436	SDRM[§]	2.5%	3.3%	Hamers et al.[226]
Burkina Faso	2003	97	IAS-USA	2.1%	4.1%	Vergne et al.[227]
Cameroon	2001	102	IAS-USA	2.9%	2.0%	Vergne et al.[227]
DRC	2008	253	SDRM[§]	3.2%	3.2%	Muwonga et al.[228]
Ethiopia	2003	92	IAS-USA	1.1%	2.2%	Kassu et al.[229]
Kenya	2008-2010	182	SDRM[§]	0.6%	0.0%	Hassan et al. [230]
Mali	2005-2006	198	IAS-USA	1.6%	1.1%	Derache et al.[231]
Mali	2010	51	SDRM[§]	3.9%	3.9%	Maiga et al.[232]
Morocco	2005-2009	82	SDRM[§]	4.9%	2.4%	Annaz et al.[233]
Senegal	1998-2007	200	SDRM[§]	2.0%	0.0%	Diop-Ndiay et al.[234]
South Africa	2008	80	SDRM[§]	1.2%	1.2%	Nwobegahay et al.[235]
Uganda	2006-2007	37	IAS-USA	0.0%	0.0%	Ndembi et al.[236]
Asia						
India	NG	49	Stanford[†]	0.0%	2.0%	Arora et al.[237]
India	2007	34	SDRM[§]	0.0%	0.0%	Chaturbhuj et al.[238]
India	2007-2009	47	SDRM[§]	0.0%	2.1%	Thorat et al.[239]
Iran	2010-2011	47	SDRM[§]	4.3%	0.0%	Jahanbakhsh et al.[240]
Israel	1999-2003	171	Stanford[†]	3.5%	3.5%	Grossman et al.[241]
Japan	1996-2006	402	IAS-USA	1.5%	2.0%	Ibe et al.[242]
Japan	2003-2008	2,573	SDRM[§]	4.3%	0.8%	Hattori et al.[243]
South Korea	2006	81	IAS-USA	1.2%	1.2%	Bang et al.[244]
Malaysia	2003-2004	100	IAS-USA[*]	0.0%	1.0%	Tee et al.[245]
Taiwan	1999-2006	786	IAS-USA	5.2%	4.1%	Chang et al.[246]
Thailand	2005-2007	151	IAS-USA	4.0%	4.0%	Apisarnthanarak et al.[247]
Thailand	2007-2010	466	SDRM[§]	1.9%	2.8%	Sungkanuparph et al.[248]
Thailand	2010-2011	330	SDRM[§]	0.6%	2.1%	Manosuthi et al. [249]

Table 2.2 (continued). Reverse transcriptase inhibitor resistance among antiretroviral-naïve individuals with HIV-1 infection of unknown duration.

Region	Years	N	Resistance Definition	Prevalence % NRTI	Prevalence % NNRTI	Reference
Asia						
Vietnam	2007-2008	47	SDRM§	0.0%	2.3%	Duc et al.[250]
Vietnam	2008-2009	92	SDRM§	6.5%	6.5%	Dean et al.[251]
Europe						
16 Countries	1996-2004	525	SDRM‡	9.3%	1.0%	Bannister et al.[252]
Belgium	2003-2006	285	SDRM‡	7.0%	3.5%	Vercauteren et al.[253]
Croatia	2006-2008	118	SDRM§	19.5%	1.7%	Grgic et al. [254]
Denmark	2001-2009	1,405	SDRM§	2.9%	1.3%	Audelin et al.[255]
Former USSR	1997-2004	278	Stanford†	2.5%	3.6%	Vazquez de Parga et al.[256]
France	2001-2002	363	IAS-USA*	4.3%	0.8%	Descamps et al.[12]
Germany	2001-2005	831	IAS-USA*	5.4%	3.0%	Sagir et al.[257]
Greece	2002-2003	101	IAS-USA	5.0%	4.0%	Paraskevis et al.[258]
Greece	2009-2011	238	SDRM§	14.3%	18.9%	Skoura et al.[259]
Italy	1996-2007	1,690	SDRM§	11.0%	6.0%	Bracciale et al.[260]
Italy	2000-2010	2,937	SDRM§	7.4%	5.0%	Colafigli et al.[206]
Italy	2001-2006	569	SDRM‡	6.5%	6.0%	Lapadula et al.[261]
Latvia	2005-2006	117	SDRM§	0.9%	0.9%	Balode et al.[262]
Romania	1997-2011	56	SDRM§	14.3%	3.6%	Temereanca et al.[207]
Slovenia	2000-2004	77	IAS-USA*	3.9%	0.0%	Babic et al.[263]
Slovenia	2005-2010	131	SDRM§	0.8%	1.5%	Lunar et al. [208]
Spain	2004-2008	683	SDRM§	4.4%	4.0%	García et al.[264]
Spain	2007-2010	1,864	SDRM§	3.9%	2.3%	Monge et al.[265]
Sweden	2003-2010	1,463	SDRM§	4.1%	2.7%	Karlsson et al.[266]
United Kingdom	1996-2004	4,138	IAS-USA*	7.8%	4.7%	UK Collaboration[210]
United Kingdom	1997-2005	8,272	SDRM‡	5.8%	3.6%	Green et al.[267]
United Kingdom	2004-2006	154	IAS-USA*	3.9%	1.3%	Booth et al.[211]
United Kingdom	2005-2006	149	IAS-USA*	3.4%	4.7%	Fox et al.[212]
United Kingdom	2005-2007	392	IAS-USA*	0.5%	1.8%	Payne et al.[268]

Table 2.2 (continued). Reverse transcriptase inhibitor resistance among antiretroviral-naïve individuals with HIV-1 infection of unknown duration.

Region	Years	N	Resistance Definition	Prevalence % NRTI	Prevalence % NNRTI	Reference
Europe						
United Kingdom	2002-2009	14,583	SDRM§	6.9%	4.1%	Dolling et al.[269]
Latin America						
Argentina	2003-2005	256	IAS-USA	1.6%	2.3%	Dilernia et al.[213]
Brazil	1998-2002	280	IAS-USA*	3.2%	1.4%	Barreto et al.[214]
Brazil	2000-2001	76	IAS-USA	1.3%	1.3%	Rodrigues et al.[270]
Brazil	2008-2009	63	SDRM§	4.8%	3.2%	Arruda et al.[271]
Brazil	2008-2009	99	SDRM§	15.0%	5.5%	Bermúdez-Aza et al.[272]
Brazil	2008-2009	82	SDRM§	5.0%	3.6%	Gräf et al.[273]
Chile	2000-2005	79	IAS-USA	0.0%	0.0%	Rios et al.[274]
Cuba	2003-2011	207	SDRM§	7.2%	1.4%	Pérez et al. [216]
DR	2007-2010	103	SDRM§	1.0%	6.8%	Myers et al.[275]
El Salvador	2008	98	SDRM§	5.1%	5.1%	Murillo et al.[217]
Honduras	2002-2003	336	SDRM‡	7.7%	7.1%	Lloyd et al.[276]
Honduras	2004-2007	200	SDRM§	3.0%	5.0%	Murillo et al.[277]
Mexico	2002-2003	96	IAS-USA*	12.5%	6.3%	Escoto-Delgadillo et al.[278]
Mexico	2005-2010	1,655	SDRM§	4.2%	1.9%	Avila-Ríos et al.
Peru	2002	326	IAS-USA	2.1%	0.6%	Lama et al.[219]
North America						
Canada	2000-2001	494	IAS-USA*	3.2%	1.2%	Jayaraman et al.[191]
USA	1999-2001	491	IAS-USA*	7.8%	3.0%	Novak et al.[192]
40 Cities	2003	317	IAS-USA	3.5%	6.0%	Ross et al.[279]
10 Cities	1997-2001	1,082	IAS-USA*	6.4%	1.7%	Weinstock et al.[194]
6 Cities	NG-2008	1,585	SDRM§	8.2%	8.3%	Poon et al.[280]
7 Cities	2006-2009	145	SDRM§	4.1%	12.4%	Hightow-Weidman et al.[281]
10 States	2007	2,030	SDRM§	7.8%	5.6%	Wheeler et al.[282]
10 States	2005-2007	228	Stanford†	4.5%	9.8%	Huang et al.[283]
Chapel Hill, NC	1999-2010	677	SDRM§	NG	4.9%	Yanik et al.[223]

Table 2.2 (continued). Reverse transcriptase inhibitor resistance among antiretroviral-naïve individuals with HIV-1 infection of unknown duration.

Region	Years	N	Resistance Definition	Prevalence % NRTI	% NNRTI	Reference
North America						
USA						
Charlotte, NC	2008-2011	189	SDRM§	2.6%	6.9%	Klibanov et al.[284]
Detroit, MI	2006-2008	133	SDRM§	9.8%	8.3%	Huaman et al.[285]
Portland, OR	2003-2009	165	SDRM§	9.1%	10.3%	MacVeigh et al. [286]
Saint Louis, MO	2003-2005	192	IAS-USA*	6.3%	7.3%	Grubb et al.[287]
Seattle, WA	2001-2009	801	SDRM§	6.0%	12.0%	Taniguchi et al.[288]
Washington, DC	2005	42	IAS-USA	2.4%	4.8%	Boyd et al.[289]
New York State	2006-2008	4.032	SDRM§	2.9%	6.3%	Redhead et al.[290]

NRTI=nucleoside(tide) reverse transcriptase inhibitor; PI=protease inhibitor; NNRTI=non-NRTI; IAS-USA=International Antiviral Society–USA Panel; SDRM=surveillance drug resistance mutations; NG=not given. *Includes 215 revertant mutations and exclude V118I mutation of HIV-1 reverse transcriptase. †Estimate includes intermediate- and high-level resistance only. ‡Based on transmitted resistance surveillance mutation list assembled by Shafer et al. [188]. §Based on surveillance drug resistance mutations list updated in 2009 [189].

Even in areas with an extensive history of widespread ARV use, there are regional differences in transmitted resistance. TDR appears to be more common in North American cohorts compared to European cohorts: surveys including 19 European cohorts over 1996-2002 reported that 11% of individuals with recent HIV infection had ≥1 major mutation [291], while in the United States and Canada, 15% of recently infected patients from 1995-2006 had evidence of TDR [221]. In a review of published literature through 2009, an estimated 12.9% (95% CI: 12.2-13.7%) of all therapy-naïve individuals in North America had evidence of any drug resistance, mostly driven by NRTI (7.4%, 95% CI: 6.8-8.0%) and NNRTI resistance (5.7%, 95% CI: 5.2-6.2%) [292]. Corresponding European studies yielded regional estimates of overall, NRTI, and NNRTI resistance of 10.9% (95% CI: 10.6-11.4%), 7.4% (95% CI: 7.1-7.7%), and 3.4% (95% CI: 3.2-3.6%), respectively [292].

Temporal trends in these regions reflect more frequent NNRTI use and an increasing proportion of patients on HAART achieving durable virologic suppression [220]. Over time, TDR declined in Europe (11.5% to 7.7% before 2001 to after 2004), driven by decreases in NRTI resistance from 8.0% to 4.3% and PI resistance from 3.3% to 1.4% before 2001 to after 2004 [292]. NNRTI resistance peaked in 2001-2002 in Europe, but declined to 3.2% after 2004 [292]. These estimates agree with results from 19 European countries, where TDR decreased from 12% to 8% from 1996-2004, and NRTI resistance decreased from 10% to 4% from 2001-2004 [193]. North America, in contrast, experienced an increase in TDR prevalence from 11.6% before 2001 to 14.3% after 2003, with increasing NNRTI resistance outpacing overall declines in transmitted NRTI and PI resistance [292].

Since TDR is influenced by ARV usage patterns, it is not surprising that NRTI resistance has dominated in North America and Europe (Table 2.1 and Table 2.2). Most

NRTI resistance detected among therapy-naïve individuals from Europe and North America are TAM including M41L, K70R, and mutations in codon 215, which are present in nearly 5% of these populations; far fewer have evidence of M184V (<1%) or K65R (<0.5%) [12, 180, 191, 192, 194, 204, 279]. NNRTI-associated mutations are usually the second most common form of TDR resistance (1-15%), and many studies report K103N, Y181C, and G190A/S. Transmitted multi-drug resistance, though, remains rare, with dual-class and triple-class resistance present in 2.6% and 1.1% of individuals infected between 2004-2005 [180].

In contrast to developed regions, TDR estimates from resource-poor settings have been traditionally nonexistent, but the WHO implemented a strategy to monitor prevalence below 2 critical thresholds (5% and 15%) in settings where resistance testing is not routine but antiretroviral therapy was on schedule for expanded access [293]. Since the rollout of antiretrovirals starting in 2003, there is evidence of an increase in TDR prevalence, but changes are highly variable across different regions. Sub-Saharan Africa, for example, experienced an estimated annual increase of 14% (95% CI: 0-29%) since rollout started in 2004, with an overall estimated prevalence of 3.7% (95% CI: 2.5-5.4%) by 2011 [294]. Eastern Africa experienced a steeper 29% annual increase (95% CI: 15-45%) since rollout, with an estimated overall prevalence of 7.4% (95% CI: 4.3-12.7%) by 2011 [294]. Transmitted NNRTI resistance dominated these trends, reflecting their use in all first line options [294]. Latin American and Caribbean countries, in contrast, experienced no statistically significant increase in TDR prevalence since an estimated 6.9% of individuals had evidence of TDR at the start of and 8 years after antiretroviral therapy access was expanded [294].

2.8.2 TRANSMISSIBILITY AND PERSISTENCE OF DRUG RESISTANCE MUTATIONS

The relatively high prevalence of TAM among therapy-naïve individuals is reflective of populations where mono- and dual-NRTI therapy as well as single PI-based HAART were once the standard of care [195], while the low prevalence of major PI mutations and some NRTI mutations may be due to a reduction in transmissibility of certain variants [295]. Transmission efficiencies of drug resistant variants relative to wild type strains have been estimated by several studies that have attempted to define populations of potential transmitters and recent seroconverters. One group estimated variants containing the M184V mutation alone were more than 10 times less likely to be transmitted than sensitive virus [296]. Dual-class and triple-class resistant variants were even less likely to be transmitted with a 20-fold reduction in transmission compared to sensitive virus [296]. Another group estimated that while the prevalence of TAM were similar between recently infected individuals and the pool of potential transmitters, M184V was underrepresented among recent infected individuals [297]. A third group estimated that M184I/V and T215Y/F had the lowest transmission efficiency relative to sensitive virus, while V118I, K219E/Q, and Y181C/I, had a higher transmission efficiency relative to sensitive virus [295]. Whether the underrepresentation of certain mutations among recently infected is due to outgrowth of more fit variants in the absence of ARV exposure [9, 13], reduction in viral load associated with these mutations among potential transmitters [297], or other factors, cannot be determined from these study designs.

Specific mutations are also differentially distributed between recently infected and chronically infected individuals, possibly reflecting persistence or reversion of these variants in the absence of ARV selective pressure. For example M184V, T215Y/F, and K103N are found more frequently among those with recent infections, while other variants not associated with phenotypic resistance, such as T215 revertant mutations, are less likely to be found in recently infected individuals [194]. Reversion cannot be

confirmed from cross-sectional comparisons between therapy-naïve populations infected at different times and by potentially different sources.

Longitudinal sampling of recent seroconverters with TDR has been applied to estimate the rate of reversion of drug resistant variants in the absence of therapy. Among 12 of 20 recently infected individuals with ≥2 pre-therapy samples available, the virus evolved at drug resistance positions without ARV pressure over a median 15 months [298]. The variants that reverted to wild type, in order of increasing time since detection, were K70R, D67N, T215S, Y181C, and M184V [298]. T215Y/F variants evolved to revertant mutations rapidly in all cases, while M41L, T69D/N, L210W, K219Q, and G190S in were maintained over follow-up [298]. These observations agree with another study of 14 recently infected individuals: while M41L, T69N, K103N, and T215 revertants persisted up to 3 years, Y181C and K219Q became undetectable in 27 months and 9 months, respectively [299]. However, each of these studies detected resistance by population sequencing, which misses minority variants that make up <20% of the total [146], making it difficult to draw conclusions about the mechanism of resistant variant decline in the population.

2.8.3 Clinical Impact of Transmitted Resistance

The general consensus is that resistant variants usually show a reduction in fitness compared to sensitive variants, as implied by *in vivo* studies of the decline of many of these mutations over time among untreated, recently infected individuals [13, 128, 299], as well as by *in vitro* comparisons of replicative capacity and infectivity between resistant and sensitive variants [10, 129, 300, 301]. It follows that TDR may be expected to have a beneficial impact on the natural course of infection. Several studies have observed lower HIV-1 RNA levels [302, 303] and higher CD4+ T cell counts [4, 303, 304] near the time of HIV infection among individuals infected with drug resistant

virus compared to individuals infected with sensitive virus. These apparent advantages may not be preserved over time, however. One study estimated that an individual with 500 CD4+ T cells/µL at seroconversion could expect a decline of 199 cells/µL or 73 cells/µL in the first year without treatment if infected with resistant or sensitive virus, respectively [304].

Once therapy is initiated, patients with TDR experience reduced virologic benefit compared to patients infected with fully susceptible virus. Among a cohort of patients on HAART for 6 months, time to suppression of HIV-1 RNA levels was longer and time to first failure was shorter for patients with high level phenotypic resistance compared to patients with sensitive virus [5]. Another study showed that patients with genotypic evidence of resistance to their initial regimen achieved suppression in a median of 3 months compared to 5 weeks for patients without resistance [4]. In another cohort study, patients with resistant HIV experienced a slower decline in plasma HIV-1 RNA between 1 and 6 months of HAART compared to patients with sensitive virus following a similar rate of decline in the first month of HAART [305]. Other studies have failed to detect statistically significant differential virologic response [17, 252, 303, 304, 306], but lack of statistical significance association could be due to the small number of exposed patients, shorter follow-up, decreased frequency of HIV-1 RNA level testing, or even definition of TDR. Most studies that reported “no effect” actually reported that fewer patients with TDR achieved virologic suppression within the study period compared to other patients. For example, virologic suppression within 6 months of HAART initiation was achieved in 85% (49/58) of patients with susceptible virus and 64% (7/11) of patients with TDR in one study ($p=0.2$) [303].

In terms of immunologic response, most studies have not demonstrated negative outcomes for patients with TDR, but these are difficult to interpret given the variable

definitions of immunologic outcomes. For example, one group reported that chronically infected patients with resistance to ≥1 ARV in their first regimen were *more* likely to achieve an increase of 100 CD4 cells/µL after 6-12 months of HAART compared to patients with >3 active ARV [252]. Similar observations were reported in an even smaller study in which patients with TDR experienced larger CD4 cell count gains at months 6 and 12 after HAART compared to patients with wild type strains [154].

2.9 MEASUREMENT OF ANTIRETROVIRAL ASSOCIATED RESISTANCE

Because a single individual infected with HIV produces 10^9 viral particles with possibly 10^6 mutations per day [307], currently available technology cannot provide a full picture of evolving HIV populations in one compartment, such as plasma, over time, let alone simultaneous characterization of viral diversity within infected resting T cells or other compartments such as the genital tract or central nervous system. Given these limitations, the standard fluid tested for antiretroviral resistance is plasma because of its relative ease of collection, storage, and processing. However, the use of plasma allows characterization of HIV-1 RNA only from actively replicating virus, not integrated HIV DNA, which may contain archived drug resistance mutations. Also, HIV-1 RNA extracted from plasma must be reverse transcribed into cDNA and then amplified via PCR prior to resistance testing to enrich for rare viral sequences, which introduces error through multiple steps and probabilistically favors amplification of the dominant circulating viral population [7, 308]. Further, individuals must have >500 HIV-1 RNA copies/mL of plasma to allow sufficient HIV RNA templates for sampling for sequencing by commercial assays [308]. Thus, commercial resistance assays sample the most recently replicating, dominant viral variants among individuals with higher viral loads.

2.9.1 GENOTYPIC ASSAYS USING POPULATION (BULK) SEQUENCING

Standard genotypic antiretroviral resistance tests (GART) rely on population or bulk sequencing, where specific regions of HIV-1 *pol* are amplified and sequenced from

a sample of HIV genomes using the chain terminator method [309, 310]. For chain terminator sequencing, 4 deoxynucleotides (dNTP) and 4 fluorescently labeled (1 color/base) dideoxynucleotides (ddNTP) are used to extend DNA strands complimentary to amplified HIV DNA (dNTP) or to terminate DNA extension (ddNTP) [309]. As a result, each sample contains fragments of HIV DNA with a labeled ddNTP at each 3' end. This population of HIV DNA fragments is separated by size, labeled fragments fluoresce as they pass through a laser light source, and the order of ddNTP is recorded on a chromatogram. The chromatogram consists of ordered, color-coded peaks for each incorporated ddNTP, representing the consensus HIV DNA nucleotide sequence. This sequence is then aligned with a reference HIV sequence to identify mutations associated with reduced ARV susceptibility.

2.9.2 INTERPRETATION OF GENOTYPIC RESISTANCE DATA

A variety of systems are available to interpret genotypic data, including proprietary systems such as the *Virtual*Phenotype [311], publicly available algorithms including the Stanford University HIV Drug Resistance (HIVdb, Appendix 2.3) [171, 184] or *Agence Nationale de Recherches sur le SIDA* (*ANRS*) algorithms [312], and lists of mutations such as those published by the IAS-USA (Appendix Table A.1) [313]. These algorithms are used to score a patient's viral sequence with a given set of mutations as susceptible, possibly resistant, or resistant to a particular ARV, or in some cases they may predict a fold-change in susceptibility for each ARV. Different algorithms may use several sources of resistance to construct rules including [221]: (1) associations between mutations and ARV exposure (genotype-treatment correlations); (2) associations between mutations and phenotypic susceptibility (genotype-phenotype correlations); and (3) associations between mutations and clinical outcomes (genotype-outcome correlations).

For genotype-treatment correlations, *in vitro* passage experiments may be used in which mutations are selected in laboratory strains of HIV in the presence of a single ARV. *In vivo* data may be used, in which sequences derived from patient isolates are correlated with ARV exposure history [130]. However, these *in vivo* correlations cannot determine whether the mutations identified are merely markers for resistance in certain populations. Further, genotype-treatment correlations determined *in vitro* identify a fraction of mutations observed in a clinical setting in which people may have complex ARV exposure histories.

Like genotype-treatment correlations, genotype-phenotype correlations may be derived from *in vitro* and *in vivo* sources. Mutations that arise from passage experiments or that are introduced into laboratory HIV strains may be tested with a phenotypic assay, or correlations between pairs of genotypic data and phenotypic data may be analyzed from patient sequence databases. Genotype-phenotype correlations may miss some clinically significant mutations such as T215 revertants, and clinical cut-offs for each ARV cannot be defined using this approach.

Unlike the previous two correlation types, genotype-outcome correlations are only estimated *in vivo*. Genotype-outcome correlations are determined using data from clinical trials in which the effects of patient baseline genotypes on virologic response to new ARV are studied. Even this approach is limited, though, since various trials have different study designs and various patient populations with different ARV experience. Thus, each data source may be weighted very differently between algorithms and produce discordant interpretations of the same set of mutations. Also, interpretation algorithms must constantly be updated to be clinically relevant as new ARV and ARV combinations are used in clinical practice.

2.9.3 GENOTYPIC ASSAYS TO DETECT MINORITY DRUG RESISTANT VARIANTS

One of the major drawbacks of commercial resistance assays is their inability to detect and quantify minority HIV variants that make up <20% of the sample [146, 308]. It is clear that resistance detected by commercial assays adversely affects treatment outcomes [4, 5, 175], but the role of undetected minority resistant variants is less clear. In a pooled analysis of 985 therapy-naïve participants who received an NNRTI-containing regimen, minority NNRTI resistant HIV was associated with an increased hazard of virologic failure (HR: 2.6, 95% CI: 1.9-3.5) [18]. This study also reported a dose-response relationship between the number of copies of reverse transcriptase inhibitor (RTI) resistant variant and virologic failure: 10-99 vs. 0 copies of resistant virus (HR: 2.2, 95% CI: 1.5-3.2); 100-999 vs. 0 copies (HR: 3.0, 95% CI: 2.0-4.5); ≥1000 vs. 0 copies (HR: 4.1, 95% CI: 2.5-6.8) [18].

Most studies included in the pooled analysis relied on allele-specific PCR (AS-PCR) to measure minority variants [127, 314, 315], where HIV DNA amplicons from an initial RT-PCR reaction are subsequently amplified by real time PCR using codon specific primers. AS-PCR (Table 2.3) is quantitative, since primers specific to the target codon are combined with HIV-1 total copy primers. As primer sets extend across targeted sequence, probes annealed within the target template sequence fluoresce, producing signals for the mutant or wild type sequence and the total copy sequence. Intensity of the signals increases with each amplification cycle, producing curves for both total copy number and mutant or wild type sequences. By comparing the curves, the relative abundance of variants in the sample may be quantified down to 0.1% [127, 315]. However AS-PCR has several major limitations: (1) Only a limited number of *a priori* mutations are interrogated.; (2) Polymorphisms within the primer or probe binding sites may lead to false negative results.; (3) RT-PCR and real time PCR may introduce bias through nucleotide misincorporation and differential amplification [19]. (4)Linkage of

mutations to the same viral genome is not possible since AS-PCR queries single codons at one time [316].

Another high throughput, low cost assay that detects minority point mutations includes the oligonucleotide ligation assay (OLA) (Table 2.3) [317]. The OLA relies on differential hybridization of mutant and wild type amplicons to specific oligonucleotide probes in a 96-well plate. Two oligonucleotide probes with different detection molecules, one for wild type, one for mutant, are added to wells along with a ligation enzyme and amplified HIV DNA. After ligation, two separate reactions are used to visualize hybridization complexes containing either wild type or mutant sequences. While OLA is more sensitive for detection of minority variants than bulk sequencing and can be performed rapidly on clinical samples, OLA is only quantitative if a standard of known relative abundance is run in parallel. Also, OLA is highly sensitive to polymorphisms in the regions queried by the probes, with incomplete binding leading to 2% of clinical samples negative for both wild type and mutant probes [317].

The heteroduplex tracking assay (HTA) is more sensitive than the OLA, allows quantification of different viral populations, and is high throughput (Table 2.3). HTA probes are generated from labeled HIV-1 DNA clones with nucleotide changes proximal to the target nucleotides associated with resistance [318-320]. These changes allow greater mismatch between the HTA probe and wild type sequences relative to mutant sequences, and secondary structure of these heteroduplexes that results from mismatches may be resolved by polyacrylamide gel electrophoresis (PAGE). Using PAGE, HTA: wild type heteroduplexes shift upward from the probe: mutant heteroduplexes, forming two distinct bands. The relative intensity of these bands allows quantification of relative abundance of variants down to 1-3% [318, 319]. Bands may also be isolated and sequenced with less contamination [320]. However, the HTA is still

limited by: (1) errors introduced by an initial RT-PCR step to reverse transcribe and amplify HIV DNA sequences; (2) poor separation between mutant and wild type heteroduplexes; and (3) restriction to predetermined resistance mutations.

All previously discussed technologies, while high throughput and low cost, rely on the detection of *a priori* resistance mutations. Other sensitive experimental technologies address this limitation, but many have been traditionally been limited by cost and labor. Clonal sequencing (Table 2.3) [321, 322], for example, is a technique in which an amplified HIV DNA insert is cloned into plasmid DNA and transformed into bacterial cells. Each bacterial colony originates from one cell carrying a single copy of the plasmid. By sequencing plasmids from multiple colonies, the relative abundance of HIV variants can be determined. However, clonal sequencing is still limited by the introduction of errors during the initial RT-PCR step [323], and by the loss of viral genomes during the cloning step. Also, there is the possibility of resampling of PCR template molecules, which can skew relative abundance estimates [19]. This last limitation has been addressed by taking a single colony from multiple PCR reactions, but this type of clonal sequencing is even more costly and labor-intensive [324].

Some limitations of clonal sequencing have been addressed by a method that relies on serial dilution of cDNA templates so that one molecule on average is amplified in each initial PCR reaction. This method, single genome sequencing (SGS) [146], employs multiple PCR reactions per cDNA dilution. PCR products are resolved on an agarose gel to determine the cDNA dilution that contains about 1 template per reaction based on the expected number of positives using the Poisson distribution. That cDNA dilution is then subjected to nested PCR, and products are bulk sequenced, allowing linkage of mutations on the same viral genome. However, sensitivity to detect minority variants is limited by the number of genomes or clones that are sequenced for both SGS

and clonal sequencing: 490 genomes/clones are required to detect a mutation present at 1% abundance with 99% certainty [316].

Alternatively, massively parallel, ultra deep sequencing (UDS) techniques, including Illumina and 454 platforms, promise to deliver the advantages of SGS and clonal sequencing in a high throughput format [22, 23, 325, 326]. For HIV applications, a library of HIV DNA must be prepared for UDS in a series of amplification, purification, and quantification steps. Typically, HIV RNA preparations from clinical samples are independently reverse-transcribed with a sequence specific cDNA primer that includes preassigned barcode nucleotide sequence [23]. These are then amplified by nested PCR, purified, quantified, and pooled for UDS library preparation.

For Roche's 454 *pyro*sequencing platform (Table 2.3) [327], two proprietary adaptors are linked to each end of HIV DNA amplicons, either through incorporation during RT-PCR or ligation after RT-PCR. Both adaptors contain a nucleotide sequence for PCR and sequencing priming, but one adaptor also contains a tag that allows enrichment of single stranded DNA linked to one of each adaptor. Single stranded HIV DNA is diluted and captured by DNA beads, one molecule per bead. Beads are combined with PCR reagents and each HIV DNA is clonally amplified on the bead by emulsion PCR [328]. Clonally amplified HIV DNA is denatured and distributed across a picoliter plate, one bead per well [329]. Nucleotides are singly washed over the wells in limited concentration. Unincorporated nucleotides are degraded, but if a nucleotide is incorporated, pyrophosphate is released, converted into ATP, and used by an enzyme to generate a light signal. The signal is recorded as a series of ordered peaks proportional to the number of identical bases incorporated.

The latest version of the 454 pyrosequencing platform, 454 GS FLX Titanium, allows for simultaneous interrogation of all known positions within 300-1,000 bp region,

producing up to 10^6 sequences per run. This enormous output allows 454 pyrosequencing to detect minority sequences present in <0.1% of the population [330]. However, multiple PCR amplification steps and the initial RT step introduce a substantial amount of error [24]. Additional error may be introduced when more than three of the same nucleotide occur in a row [24], resulting in false calls of the third nucleotide, followed by partial incorporation of the next base [325]. These errors can make it difficult to distinguish actual base changes associated with drug resistance from errors introduced by reverse transcription or PCR amplification.

Illumina UDS platforms, including MiSeq and HiSeq 2500 (Table 2.3), offer an alternative to the 454 sequencing platform. Similar to the 454 platform, proprietary adaptors are added to the ends of double stranded amplified HIV DNA to enrich single-stranded HIV DNA with both adaptors. These templates are then immobilized on a surface coated with sequences complimentary adaptors [326]. Immobilized HIV DNA is clonally amplified by bridge PCR to form dense clusters of HIV DNA. One strand of the HIV DNA template is sequenced at a time, depending on which adaptor is cleaved prior to denaturation of amplified, double stranded HIV DNA products [326]. After sequencing the initial strand, the immobilized HIV DNA is again amplified, and double stranded DNA is nicked at the opposite end for paired end sequencing.

For the sequencing using the Illumina platform, DNA polymerase and four labeled nucleotides are added to the surface containing clusters of clonally amplified, single stranded HIV DNA. The DNA polymerase incorporates only one nucleotide, since the nucleotide's 3' hydroxyl group of the nucleotide is blocked after its incorporation [331]. Unincorporated nucleotides are washed away, and the color of the incorporated nucleotide is captured for each immobilized HIV DNA cluster. A second enzyme then

cleaves the blocking group from the incorporated nucleotide, and these molecules are washed away following cleavage, allowing another single round of sequencing to begin.

Compared to 454 platforms, the Illumina sequencing platforms are higher output, producing up to 6 billion 150 bp paired end sequences per run for HiSeq 2500 and 15 million 250 bp paired end sequences for MiSeq. Illumina platforms are less prone to errors within homopolymeric stretches of sequence compared to 454 technologies [332], but the quality of the sequencing read declines rapidly towards the end of the run [333]. The decline in sequence quality over the run is due to incomplete cleavage of the nucleotide blocking site that accumulate over the run [334]. Also, low diversity sequences, such as might be expected among a pool of amplified, HIV DNA with predetermined barcodes, cause the Illumina platform data filters to count multiple clusters as the same sequence and bias estimates of sequence diversity [335].

Table 2.3. Common genotypic assays to detect low abundance antiretroviral resistant HIV-1.

Assay	Avg. Length	Threshold	Main error source	Cost	Throughput
AS-PCR	point mutation	0.1%	primer/probe mismatches	$$	high
OLA	point mutation	10%	sequence mismatches	$	high
HTA	point mutation	1-3%	band discrimination	$$	medium
Clonal	700 bp	5%*	cloning efficiency	$$$	low
SGS	700 bp	5%*	dilution	$$$	low
454	700 bp	<1%	homopolymeric regions	$$$$	high
Illumina	150-250 bp	<1%	phasing	$$$$	very high

AS-PCR=allele-specific polymerase chain reaction; OLA=oligonucleotide ligation assay; HTA=heteroduplex tracking assay; SGS=single genome sequencing. *Value for sequencing 60 genomes with 95% certainty; sequencing 490 genomes is necessary to obtain a 1% limit with 99% certainty.

2.9.4 ADDRESSING MULTIPLE SOURCES OF SEQUENCING ERROR

Despite rapidly advancing next generation sequencing techniques, estimating resistant HIV populations in plasma remains challenging because of the balance between the high diversity of rapidly evolving HIV within the individual [21] and errors introduced during multiple steps of viral enrichment and sequencing. These errors are first a function of the integrity of starting samples, which must be properly handled and

stored to prevent loss of HIV RNA genomes [332]. Upon RNA extraction from these samples, genomes are inevitably lost, but various protocols may be used to enhance recovery. For example, plasma samples with low viral titers may be centrifuged prior to extraction to concentrate viral particles [333], or carrier RNA may be added to samples so that rare viral RNA sequences are less likely to be lost during extraction.

Even if losses during RNA extraction are minimized, errors may be introduced when purified HIV RNA is reverse-transcribed into cDNA by an engineered RT enzyme, which introduces an error every 30,000 to 17,000 nucleotides [334]. Reverse transcription is also associated with template recombination, which occurs when cDNA strands are incompletely extended and serve as primers for downstream amplification. Reverse transcription errors may first be addressed by starting with the highest quality (full-length) RNA, and by increasing cDNA synthesis reaction times to allow complete extension of the cDNA [335].

The next step, PCR amplification, is a major source of sequencing error, but it is currently necessary to produce an adequate sequencing signal. PCR errors may be introduced through one or more of the following: (1) biased amplification due to mismatches between the primer and some HIV sequences [336]; (2) resampling of starting templates due to low number of input templates [19]; (3) nucleotide misincorporation by the DNA polymerase [337]; and (4) PCR recombination from incomplete extension of amplified DNA or template switching [338]. To alleviate biased amplification due to primer-template mismatches (1), primers may be designed to complement conserved sequence regions or degenerate nucleotides may be substituted where potential mismatches may occur. To address PCR resampling (2), the number of starting templates may be maximized by concentrating the virus before extraction, for example. To reduce nucleotide misincorporation (3), high fidelity DNA polymerases with

proofreading capability are often used, but there is usually a trade-off with PCR efficiency [337]. Finally, to prevent PCR recombination (4), PCR extension times must be optimized to prevent too-short products from priming subsequent amplifications or to reduce the opportunity for template switching by the DNA polymerase [338].

For next-generation sequencing platforms, errors introduced during PCR and some platform specific errors discussed previously (2.9.3) have been addressed using a method engineered into the cDNA synthesis step. Here, a randomized sequence (Primer ID) is tagged to each individual viral genome via the cDNA synthesis primer (Figure 2.5A) [26]. The Primer ID and pre-assigned, sample specific nucleotide barcode are subsequently amplified along with the viral sequence (Figure 2.5B). Downstream, majority-rules consensus sequences may be constructed from 3 or more resampled sequences within a sample (barcode) sharing the same Primer ID sequence (Figure 2.5C), so that concordant nucleotides at a given position are more likely to be genuine nucleotide calls and discordant nucleotides at a given position are more likely to be errors. Importantly, collapsing raw sequences into Primer ID consensus sequences allows estimation of the viral population that was originally sampled, rather than what was best amplified. Thus, tagging individual viral genomes with a Primer ID allows correction of most sources of error from PCR amplification forward.

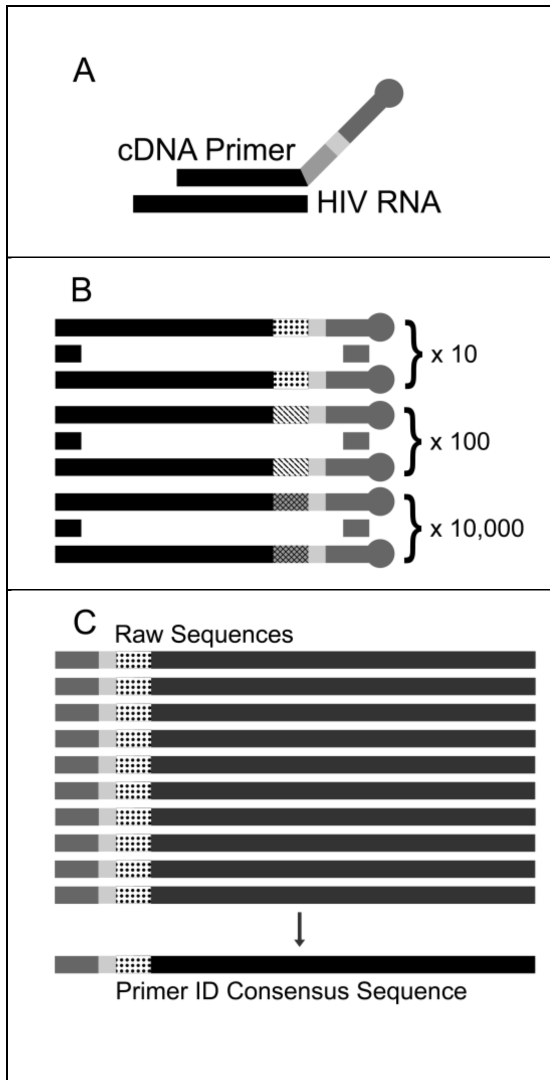

Figure 2.5. Primer ID method to estimate HIV-1 viral population [26]. Panel A demonstrates how a randomized nucleotide sequence (Primer ID, medium gray) is added to the HIV RNA genome (black) during cDNA synthesis. The cDNA primer consists of an HIV specific sequence (black), the sequence specific Primer ID (medium gray), the pre-assigned sample barcode (light gray), and the non-specific sequence for downstream PCR (dark gray with circle). Panel B shows differential amplification of different Primer ID (black circles, diagonal stripes, dark gray hatch on gray) tagged viral genomes within the same sample (light gray barcode). Panel C shows how multiple resampled raw sequences (dark gray) sharing the same Primer ID (black circles) and barcode (light gray) are collapsed into a single consensus sequence (black).

CHAPTER 3 RESEARCH DESIGN AND METHODS

3.1 OVERVIEW

For this study, a new method (Primer ID) that simultaneously addresses many sources of PCR and ultra deep sequencing (UDS) errors [26] was used to estimate HIV-1 reverse transcriptase (RT) inhibitor resistance among a population of therapy-naïve, HIV-infected patients who eventually received an RT inhibitor regimen (N=184). For the Primer ID method, each viral RNA template was tagged with a unique primer identification sequence prior to amplification (Figure 2.5). Downstream, consensus sequences made from sequences sharing the same Primer ID represented the viral templates input into cDNA synthesis, rather than what was most efficiently amplified.

First, by sequencing known HIV-1 RT sequences spanning HIV-1 RT codons 34-138 with the 454 and Illumina deep sequencing platforms, the amount of error associated with the Primer ID was quantified for each platform (Aim 1). The rate of nucleotide substitutions, insertions and deletions, and template recombination was estimated from both Primer ID consensus and raw sequences and compared by analysis method (Primer ID vs. raw sequences) and platform (454 vs. Illumina).

Next, the 454 platform was used to sequence Primer ID tagged, amplified HIV-1 DNA spanning HIV-1 RT codons 34-138 and 139-245 from 184 therapy-naïve patients to estimate the prevalence of RT inhibitor resistance (Aim 2A). Prevalence estimates obtained using 454 deep sequencing and Primer ID were compared to those obtained by bulk sequencing. For a subset of patients with evidence of increased sequence sampling depth using the 454 platform (N=19), samples were also submitted for paired-end

sequencing of HIV-1 RT codons 34-74 and 111-138 using the Illumina MiSeq platform. For these 19 patients, results from the analysis of Illumina MiSeq sequences were compared to results obtained using the 454 platform within the same HIV-1 RT coding region (Aim 2B).

Finally, the effect of pre-existing RT inhibitor resistance on time-to-first virologic failure was estimated among 184 patients using both deep and bulk sequencing measured pre-existing resistance (Aim 3). To define pre-existing resistance, the number of active antiretrovirals within each patient's initial RT inhibitor regimen was determined using the Stanford HIV Drug Resistance Database (HIVdb) [171] to interpret Primer ID consensus sequences and sequences obtained using bulk sequence analyses (See Appendix 2.3 for Stanford HIVdb scores.). Effect-measure estimates were obtained for deep sequencing alone, for bulk sequencing alone, and for the combined effect of resistance detected using both technologies to determine the added value of more sensitive sequencing technology.

3.2 Study Population

Patients included in this study (N=184) were selected from patients enrolled in the University of North Carolina Center for AIDS Research HIV Clinical Cohort Study (UCHCC), an ongoing clinical cohort study founded in January 2000 that enrolls adults aged 18 years or older that are receiving HIV care at the University of North Carolina (UNC) Infectious Disease (ID) Clinic [339]. Patients enrolled in the UCHCC are representative of the HIV epidemic in the Southeastern United States, where HIV disproportionately affects minorities, the mode of transmission is primarily sexual, and most patients enter care during the chronic phase of HIV infection [340-342]. Through 2011, 59% of all patients enrolled in the UCHCC (N=3,141 total) were African American, 30% were female; 28% reported men who have sex with men (MSM) as a primary HIV

risk category, while only 11% reported injection drug use (IDU) [343]. Among 853 UCHCC patients initiating HIV care at the UNC ID Clinic between 1999-2009, the median CD4+ T cell count at entry was 286 cells/μL (inter-quartile range (IQR): 63-482) [342]. Compared to the rest of the UCHCC cohort, this sample of 184 patients included in this study were less likely to be female (23%) or African American (53%), and more likely to be MSM (44%). These 184 patients were, however, enrolled in the cohort during the chronic stage of HIV infection, with a median CD4+ T cell count of 307 cells/μL (IQR: 123-444) at entry to care.

Only a subset of UCHCC patients was eligible for this study, based on a set of criteria related to the availability of a clinical sample for resistance testing and the timing and content of the initial antiretroviral regimen. First, UCHCC patients consented to the use of their specimens, which were collected as part of routine clinical care. Second, patients initiated therapy after December 31, 1999 to alleviate potential survivor bias from including patients who initiated therapy prior to the establishment of the cohort. Third, patients initiated therapy with a HAART regimen composed exclusively of RT inhibitors, including either ≥3 nucleoside(tide) reverse transcriptase inhibitors (NRTI), or ≥2 NRTI and 1 non-NRTI (NNRTI). This criterion was imposed since sequencing was restricted to the coding region of HIV-1 RT associated with RT inhibitor resistance. Finally, patients had ≥1 pre-therapy HIV-1 RNA level recorded in the database, indicating a potential available sample for sequencing. A total of 331 UCHCC patients met the above criteria as of August 17, 2012, and of these, 184 had an archived, pre-therapy plasma sample available for sequencing. A flow diagram of the study population is shown in Figure 3.1.

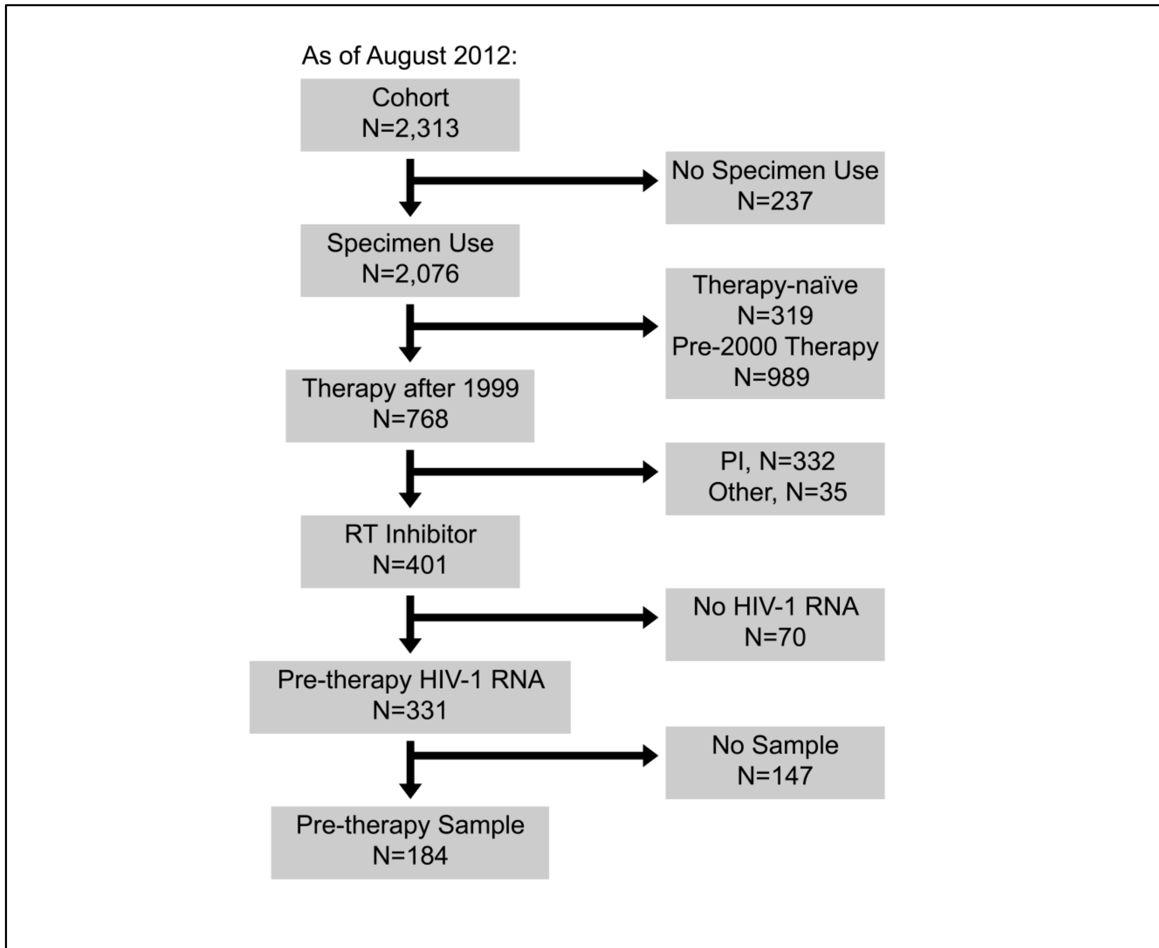

Figure 3.1. Flow diagram of the study population. RT inhibitor regimens consisted of ≥3 nucleoside(tide) reverse transcriptase inhibitors (NRTI) or ≥2 NRTI plus 1 non-NRTI (NNRTI). PI=protease inhibitor; RT=reverse transcriptase.

3.3 THE UNIVERSITY OF NORTH CAROLINA CENTER FOR AIDS RESEARCH

This study relied on data collected on behalf of the University of North Carolina at Chapel Hill Center for AIDS Research (UNC CFAR). UNC CFAR is a collaborative institution among investigators at the UNC School of Medicine, Family Health International, and Research Triangle Institute [344]. UNC CFAR facilitates clinical care of persons living with HIV disease in the Southeastern United States, provides support for ongoing HIV/AIDS research, and supports grant development for new HIV/AIDS research.

The UNC CFAR oversees and maintains the UCHCC, which was established in 2000 by the CFAR Clinical Core, but includes patient records dating back to the early years of the UNC ID Clinic. UCHCC maintains both an electronic database containing patient data and a repository of patient samples collected during routine care. The UCHCC database consists of patient data from standardized, biannual chart abstractions and from a hospital-wide electronic database that is updated in real time. UCHCC data is stored on a protected warehouse server with controlled access. Fewer than 2% refuse to participate in the UCHCC, and analyses indicate that patients who refuse are demographically and clinically similar to patients who agree to be in the UCHCC.

3.4 Data Sources

Medical charts for patients enrolled in the UCHCC are routinely abstracted into the database at enrollment and at six month intervals. Copies of medical charts and web based clinic notes are used to complete paper chart abstraction forms, which are then entered into the electronic database using a standardized electronic data entry form. Chart abstraction forms are the source of the following information: patient name; unique identifier for UCHCC (patient key); HIV diagnosis history; entry to HIV care; clinic visit dates; HIV risk factors; allergies; results of screening tests (chest X-rays, PAP smears, mammograms); immunizations; hospitalizations; illnesses; medication history, including antiretroviral regimens and other medications; and HIV-related lab results not obtained at UNC. Paper forms are secured in a locked cabinet until they are completely entered into the UCHCC database, after which they are destroyed. The electronic UCHCC database is password protected and maintained on a secure warehouse server.

As previously stated, the UCHCC database also contains data that is electronically transferred from UNC Hospital's electronic database. This institutional

database is updated in real time and includes demographic information and laboratory results from tests performed through UNC Hospitals (e.g. HIV-1 RNA level and CD4 cell count). Both this electronically-transferred data and the manually abstracted data that are housed in the UCHCC were the source of clinical data for the 184 patients included in this study. Specific variables and specimens used are listed in Table 3.1.

Table 3.1. Variables and specimens from the University of North Carolina Center for AIDS Research HIV Clinical Cohort Study

Variable/clinical sample	Source
patient key*	chart review
first/last name*	UNC Hospitals
medical record number*	UNC Hospitals
date of entry to HIV care	chart review
consent questionnaire answers/date of enrollment	chart review
clinic visit dates	chart review
date of death	chart review
resistance test results/test dates†	chart review
HIV-1 RNA levels/test dates	UNC Hospitals
CD4+ T cell counts/test dates	UNC Hospitals
antiretrovirals used/start and stop dates	chart review
date of birth	UNC Hospitals
gender	UNC Hospitals
race/Latino ethnicity	UNC Hospitals
HIV risk category	chart review
≥140 μL plasma/draw date‡	repository

***Medical record numbers, names, and patient key were linked to locate archived plasma samples. †Bulk sequencing results were provided at the amino acid level. ‡All plasma samples were collected prior to the initiation of antiretroviral therapy. If possible, the sample drawn closest to the last pre-therapy bulk sequencing result was used.**

3.5 HIV-RELATED LABORATORY PROCEDURES

3.5.1 BLOOD SPECIMEN COLLECTION AND PREPARATION

For UCHCC patients, venous blood is drawn as part of routine care and stored in the repository following informed consent. Specimens retrospectively tested for resistance in this study were collected from participants prior to initiating HAART (N=184) (Figure 3.1). On average, 2 samples consisting of 2-4 mL of blood each are drawn per patient prior to initiating antiretroviral therapy (IQR: 1-4), but for these 184 patients, the median number of blood draws was 3 (IQR: 2-6). Blood samples are

centrifuged within 24 hours of collection to separate plasma from whole blood. Plasma is then used for testing or stored by the UNC CFAR Retrovirology or Immunology Core.

Upon receipt of plasma, UNC CFAR personnel update a master specimen log with barcode information, and they create a tracking sheet for each specimen. Tracking sheets and specimen vials contain the patient's name, UNC Hospital system medical record number, barcode, the sample's collection date, batch number, and protocol number. Tracking sheets are updated with each freeze/thaw per sample as repeated cycles threaten the integrity and purity of clinical specimens. To prevent multiple cycles, stored plasma is usually aliquoted into smaller volumes (1 mL).

For this study, a minimum of 140 µL plasma from each of 184 patients was used for resistance testing. For samples with fewer than 4.5 $\log_{10}$ HIV-1 RNA copies/mL, 1 mL of plasma was centrifuged (20,000 x g) for 2.5 hours at 4°C to concentrate the virus; this cut-off was chosen based on early tests of PCR primers (data not shown). After the centrifugation step, the supernatant was removed and stored at -80°C, and remaining virus-containing pellets were resuspended in 140 µL of phosphate-buffered saline (PBS) for RNA extraction.

3.5.2 HIV-1 RNA Extraction

Total RNA was extracted from plasma samples to isolate RNA from impurities and nucleases that may degrade RNA. A total of 140 µL of plasma or PBS suspension was necessary to produce 60 µL of total RNA using the QIAGEN extraction protocol (QIAGEN, Hilden, Germany); however, high HIV-1 RNA viral loads were ideal (>4.5 $\log_{10}$ copies/mL) for better sampling of the viral population. Given the instability of RNA [332, 345] and the nature of researching rare HIV sequence variants, plasma samples and RNA were handled in an area where the potential for contamination and RNA degradation was minimized. A separate clean room free of amplified/cloned DNA,

enclosed safety cabinets with laminar air flow to prevent contamination, and ribonuclease (RNase) free surfaces and consumable items were employed to protect the integrity of patient samples.

From each thawed plasma sample or PBS suspension, 140 µL was removed and combined with 5.6 µg of carrier RNA (to prevent template loss and aid RNA precipitation) and lysis buffer to: (1) denature RNases naturally present in plasma; (2) disrupt HIV-1 envelope and capsid for RNA isolation; and (3) facilitate RNA capture. Following lysis, RNA was precipitated in ethanol, and applied to a silica membrane to bind nucleic acids. Bound RNA was washed with ethanol and high salt solutions, dried to remove excess ethanol, and dissolved (eluted) from the membrane using 60 µL of RNase-free water. Purified RNA preparations were stored at -80°C in a locked freezer.

3.5.3 HIV-1 cDNA Synthesis

Next, HIV-1 RNA must be reverse transcribed into complimentary DNA (cDNA) using reverse transcriptase, since most downstream analyses rely on a DNA template. First, purified total RNA and dNTP were combined with an oligonucleotide primer such as a oligo(dT) which binds the HIV-1 RNA poly(A) tail, or a random typically 6-nucleotide sequence which could potentially bind any site within the genome, or HIV-specific nucleotide sequence that targets a specific sequence. This solution was heated and rapidly cooled to 2-4°C to disrupt the secondary structure of RNA and allow the primer to anneal to the RNA template. Next, an RT enzyme, RNase inhibitors, and reaction buffer were added, and the reaction was heated to allow cDNA extension. Following extension, the reverse transcriptase was inactivated, and RNase H was added to degrade remaining RNA bound to cDNA. cDNA may was then stored at -20°C.

cDNA synthesis was performed in two separate reactions for each sample (184 samples x 2 = 368 reactions), one per cDNA primer (Appendix 3.1). Each 60 µL cDNA

reaction contained 20 μL total RNA, 600 Units of SuperScript III (Life Technologies), 1st Strand Synthesis Buffer (Life Technologies), 0.25 mM dNTP mix, 100 mM dithiothreitol, 120 Units of RNase OUT (Life Technologies), and 0.25 μM primer specific to either HIV-1 HXB2 nucleotides 2965-2992 or 3258-3284 (Appendix 3.1) [346]. In addition to HIV-specific sequence, cDNA primers featured a random 8-nucleotide Primer ID sequence (CGNNNNNNNNTC), a 4-nucleotide sample barcode sequence (BBBB), and a non-HIV-specific PCR primer sequence (GCCTTGCCAGCACGCTCACAGCTGGCA). cDNA primers were removed following synthesis to prevent the excess from serving as primers for multiple sequences during downstream amplification, thus defeating the purpose of Primer ID. cDNA reactions were purified using the PureLink PCR Purification Kit (Life Technologies, Carlsbad, CA) with four wash steps to ensure their complete removal [26]. This kit includes a high cut-off DNA binding buffer that contains isopropanol, allowing separation of higher molecular weight cDNA from shorter cDNA primer sequences.

3.5.4 HIV-1 DNA Amplification

Non-quantitative PCR was used to amplify HIV-1 cDNA spanning HIV-1 *pol* (HXB2 nucleotides 2620-2992 and 2992-3284). For the 1st round of PCR, 500 copies of cDNA, estimated using HIV-1 RNA level of the sample, were input into a 50 μL PCR reaction (one reaction/HIV-1 fragment) containing 0.5 μM of each round 1 PCR primer listed in Appendix 3.1. In addition, each reaction contained 0.2 mM of dNTP mix, 1.25 Units of Phusion® High Fidelity Hot Start II DNA polymerase (Thermo Scientific, Waltham, MA), and Phusion® High Fidelity PCR Buffer with 1.5 mM $MgCl_2$ (Thermo Scientific). cDNA were amplified using the following parameters: (1) 30 seconds at 98°C; (2) 30 cycles of 10 seconds at 98°C, 30 seconds at 67°C (HXB2 nucleotides 2620-2992) or 63°C (HXB2 nucleotides 2992-3284), 72°C for 30 seconds; and (3) final extension for 10 minutes at 72°C. Next, 1-2 μL of the 1st PCR reaction was added to 25 μL containing identical reagents, but with 2nd round PCR primers listed in Appendix 3.1. Reactions

were amplified over 30 cycles using previous parameters, except that for the first 10 amplification cycles, annealing temperatures were lowered to 57°C and 53°C, respectively.

Given the large number of clinical samples and total reactions, cross-contamination between patient samples was a significant concern. Multiple efforts were made to reduce the possibility of cross-contamination, including using biosafety cabinets, aliquoting reagents for single use, using disposable consumables, and limiting the number of samples processed per day. A negative control was included in the first and second rounds of PCR amplification to monitor contamination. For the 1st round of PCR, water was added instead of the cDNA template (no template control), and 2 µL of this amplified product was added to the 2nd round PCR reaction to check for low level contamination. In addition, separate water-only, no template controls were set up during the 2nd round of amplification to help determine the source of contamination.

Following amplification, 10% (2.5 µL) of each sample or control was removed for agarose gel electrophoresis to visualize the amplified DNA band. The upstream HIV-1 RT amplicon was expected produce a 410 bp band including 373 bp of HIV-1 sequence (HXB2 nucleotides 2620-2992) and 37 bp of non-HIV-1 primer sequence, while the downstream amplicon was expected to produce a 357 bp band, including 320 bp of HIV-1 sequence (HXB2 nucleotides 2992-3284) and 37 bp of non-HIV-1 primer sequence. For each sample, band intensity for each amplicon was compared to a DNA mass standard to determine its concentration (Carestream Molecular Imaging Software SE, Rochester, NY). Samples were then pooled in equimolar amounts for UDS based on the determined concentration.

Pools of amplicons from multiple samples were digested with *PvuII* (recognition sequence CAG▼CTG) to cleave a portion of the non-HIV-1 PCR primer sequence from

each amplicon, thus increasing the read length of HIV-1 sequence. It is possible that some amplicons would have the *PvuII* recognition site within the Primer ID sequence and therefore be excluded from downstream analyses after cleavage; however, this would only occur in 3 different Primer ID patterns: NCAGCTGN, NNCAGCTG, and CAGCTGNN. Thus, $4^2+4^2+4^2=48$ out of a total $4^8=65,538$ possible Primer ID would contain a *PvuII* cut site (0.073%). Following digestion, pools of amplicons were separated by agarose gel electrophoresis and gel purified to remove any short sequences (QIAquick Gel Purification kit, QIAGEN). Amplicons submitted for standard bulk sequencing analysis were not pooled, but rather 2nd round PCR products were individually purified before sequencing (QIAamp PCR Purification kit, QIAGEN). Both purified individual amplicons and amplicon pools were stored at -20°C.

3.5.5 BULK HIV-1 DNA SEQUENCING

Most patients (N=180/184, 98%) had a standard genotypic resistance result for at least one sample drawn prior to initiating therapy. For the majority of patients (N=140/180, 78%), genotypic resistance testing was performed using HIV GenoSure™ or GenoSure™ *Plus* assays (LabCorp, Research Triangle Park, NC), which produce a summary report of amino acid changes over HIV-1 RT codons 20-399 (in addition to protease and integrase). For 103 (74%) of these patients, resistance genotypes were preformed retrospectively as part of a collaboration with Virco laboratories (Mechelen, Belgium). For the other 37 (26%) patients, resistance tests were performed as part of routine care to guide initial therapy selection.

For patients without commercial genotypic results (N=44), in-house bulk sequencing of both strands of each HIV-1 RT amplicon was attempted using purified 2nd round PCR products (with barcode and Primer ID) and 10 pmoles each of the inner PCR primers (Appendix 3.1). For 40 patients (91%), an in-house bulk sequence was obtained

for at least one of the two amplicons; for 8 of 40 patients, the downstream RT amplicon was not successfully sequenced.

DNA amplicons were sequenced by the UNC-CH Genome Analysis Facility on an Applied Biosystems (ABI) 3730x1 DNA Analyzer (Life Technologies) using the ABI PRISM™ BigDye™ version 1.1 Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq® DNA polymerase FS (Life Technologies). Sequence chromatograms of both strands of each HIV-1 RT amplicon were inspected for quality, trimmed, and assembled using Sequencher version 4.8 (Gene Codes Corporation, Ann Arbor, MI). To check for evidence of cross-contamination between patient samples, sequences were aligned using a ClustalW2 server [347]. Alignments were input into MEGA version 5.2 to construct Neighbor-Joining phylogenetic trees using 1,000 bootstrap replicates [348]. The Kimura 2-parameter method was used to estimate evolutionary distances between sequences [349]. Nucleotide sequences were translated into amino acid sequences and antiretroviral resistance was interpreted using the Stanford HIV Drug Resistance database (Stanford HIVdb) [171].

3.5.6 ULTRA DEEP HIV-1 DNA SEQUENCING

Ultra deep sequencing (UDS) of amplicons from clinical samples was performed using two different proprietary platforms, Roche's 454 GS FLX (Roche Diagnostics, Basel, Switzerland) and Illumina's MiSeq (150 bp paired-end, Illumina, San Diego, CA), which were discussed in previous sections. For UDS using the 454 platform, a pool of amplicons from 184 patients spanning both RT fragments was submitted for sequencing on an entire plate (10^6 sequences expected). For Illumina MiSeq, amplicons from 19 patients were sequenced using the 150 bp paired-end sequencing protocol to sequence HXB2 nucleotides 2648-2770 and 2878-2964 (10^7 sequences expected). These patients were selected because they had the greatest number of Primer ID consensus

sequences generated from 454 deep sequencing data, which is evidence of greater sampling depth of the viral population.

3.5.7 HIV-1 RNA Level and CD4+ T Cell Count

HIV-1 RNA level was measured using a quantitative RT-PCR system developed by Roche Diagnostics Systems (Basel, Switzerland). The Roche Amplicor HIV-1 Monitor assay relies on the following steps to quantify the amount of HIV-1 RNA in patient plasma: (1) plasma separation and HIV-1 RNA extraction; (2) reverse transcription of HIV-1 RNA into cDNA using biotinylated primers for a labeling reaction in steps 4 and 5; (3) PCR amplification of the cDNA; (4) hybridization of the PCR amplicons to immobilized HIV-specific probes; and (5) laser detection of the fluorescence from biotinyl-labeled primers. The Roche Amplicor HIV-1 Monitor assay can be performed using the standard or ultrasensitive procedures: the standard procedure can quantify HIV-1 RNA over a range from 400-750,000 copies/mL; the ultrasensitive procedure can quantify HIV-1 RNA over a range of 50-75,000 copies/mL.

CD4+ T cell counts were measured in-house using single-platform technology and flow cytometry. This technology is capable of measuring CD4+ T cell concentrations as low as 10 cells/μL of plasma. Normally, CD4+ T cell counts range from 500 to 1500 cells/μL, but many advanced HIV patients have CD4+ T cell counts below 200 cells/μL.

3.5.8 Plasmid Control Sequence

To characterize errors associated with both Illumina and Roche UDS platforms, all of HIV-1 reverse transcriptase (HXB2 2550-3515) from a patient HIV-1 subtype C isolate was cloned into a plasmid vector, pcDNA3.1 (Life Technologies) and used as a known sequence control. The sequence of the plasmid insert was confirmed by in-house bulk sequencing. Once confirmed by bulk sequencing, the plasmid was linearized by *BamHI* digestion, purified using the Minelute PCR Purification kit (QIAGEN), and

quantified by UV spectrophotometry. Control DNA was then serially diluted to solutions containing 3,000, 10,000, 30,000, and 100,000 copies. Double stranded plasmid DNA were denatured, cooled, and tagged with Primer ID and barcode (one per dilution) using Platinum Taq DNA polymerase (Life Technologies). Excess cDNA primers were removed [26], and HIV-1 HXB2 nucleotides 2620-2992 were amplified from each dilution using the same conditions as patient samples. Amplicons were pooled across copy number dilutions in equimolar concentration, and pools were sequenced over HIV-1 RT codons 34-139 using the 454 Junior (HXB2 nucleotides 2648-2964, Roche) and Illumina MiSeq sequencing platforms (HXB2 nucleotides 2648-2840 and 2782-2964).

3.5.9 Deep Sequencing Data Pipeline

Deep sequencing data was processed using a custom pipeline of filters, implemented slightly differently depending on the sequencing platform. Given the enormous amount of data output by Illumina MiSeq, a sample of 1,000,000 sequences was input into the filtering pipeline. For both platforms, sequence read lengths were examined and reads <300 nucleotides were discarded. Next, sequences were compared to HXB2 *pol* to determine orientation and location. The barcode and Primer ID of these sequences was checked, and sequences with invalid barcodes or Primer ID were discarded. Remaining sequences were partitioned by barcode (sample) and then Primer ID (viral template). For sequences read by the 454 platform, sequences with a Primer ID occurring fewer than 3 times within a sample were discarded; for the Illumina MiSeq platform, sequences with a Primer ID occurring less than 5 times were discarded. A larger threshold was chosen for sequencing data from the Illumina platform since a larger number of sequences were available. Consensus sequences were then created from multiple sequences with identical Primer ID. These nucleotide sequences were then translated into amino acid sequences and resistance to particular antiretrovirals was predicted using the Stanford HIVdb [171]. Excluding the HIV-specific primer

sequences used in PCR amplification, UDS sequences spanned HXB2 nucleotides 2648-2964 and 2993-3257 (454 FLX sequencing platform only).

Primer ID must be correctly sequenced to accurately estimate the viral population within the individual patient. Since each Primer ID sequence is 8 nucleotides, and since errors begin to occur with greater frequency after 3 or more identical consecutive nucleotides, homopolymer-associated miscalls within the Primer ID sequence were a concern. Miscalls within the Primer ID sequence could artificially inflate the number of templates within a sample when consensus sequences are generated from unique Primer ID. If a homopolymeric Primer ID appears a vast number of times within a sample (resampled), some small portion of Primer ID may consistently be misread and separate consensus sequences may be created from these unique Primer ID. To check for this type of error, Primer ID clusters were examined within patient barcodes with evidence of resampled Primer ID (>500 occurrences of the same Primer ID) using multiple sequence alignment. If Primer ID was 1 or 2 nucleotides different from larger clusters (resampled Primer ID), these consensus sequences were discarded.

3.6 Statistical Analyses

All statistical analyses were performed using SAS version 9.3 (SAS Corporation, Cary, NC) and Stata version 10 (StataCorp, College Station, TX).

3.6.1 Specific Aim 1

Using a known HIV-1 reverse transcriptase sequence, define a baseline error rate for the 454 and MiSeq ultra deep sequencing platforms. For each platform, compare results estimated using Primer ID consensus sequences to those estimated using raw sequences.

Known HIV-1 Reverse Transcriptase Sequence: The plasmid control for this study was created by co-contributor Shuntai Zhou, using an HIV-1 subtype C sequence

isolated from a patient sample. The entire length of HIV-1 reverse transcriptase was cloned into a pcDNA3.1 plasmid DNA vector. Plasmid DNA was quantified and serially diluted into fixed volumes containing 3,000, 10,000, 30,000, or 100,000 copies of the HIV-1 sequence insert. Each copy number dilution was input into four independent, single round DNA synthesis reactions, where DNA templates were tagged with a barcode and random 8-nucleotide Primer ID. Excess cDNA primers were removed and samples were independently amplified by semi-nested PCR. Amplified DNA was pooled across dilutions (barcodes) in equal amounts and submitted for deep sequencing using the Illumina MiSeq and Roche 454 Junior platforms.

Homopolymeric Regions: One of the main problems with 454 technologies is the tendency to read too few or too many nucleotides after ≥5 of them occur consecutively within a sequence [24, 350]. This is because one nucleotide is added to the sequencing reaction at one time, and the signal generated is proportional to the number of identical nucleotides incorporated in a row; however, the variability in the relationship between the size of signal peak and the number of identical nucleotides in a row rapidly increases as the length of the homopolymer increases, resulting in erroneous nucleotide calls. Homopolymeric regions are abundant within the interrogated region of HIV-1 reverse transcriptase (≈20-25% of the sequence), and homopolymer-associated errors were expected to be common among HIV-1 sequences read by this platform.

In this analysis, homopolymeric regions were defined as 4 or more identical nucleotides in a row plus the two flanking nucleotides. By this definition, 75 (24%) of 317 nucleotides spanning HXB2 2648-2964 were expected to be susceptible to homopolymeric error for the HIV-1 subtype C control sequence. Since error estimates from control experiments were extrapolated to subtype B patient samples, homopolymer-associated positions were defined separately for patient samples. The

2004 subtype B consensus sequence, available from the Los Alamos HIV sequence database, was used to identify homopolymeric positions for clinical patient samples since the true sequences are unknown. Within the 2004 HIV-1 subtype B consensus sequence, 61 (19%) of 317 positions were near or within homopolymeric regions within the sequence spanning HXB2 2648-2964, while 45 (17%) of 265 positions spanning HXB2 2993-3257 were expected to be within homopolymeric influence.

***Analysis*:** The main objective of Aim 1 was to define a baseline error rate for the Primer ID method using a known sequence to provide context for resistance estimates within clinical samples, where the viral population is unknown. For each sequencing platform, raw sequences or Primer ID consensus sequences were compared to the sequence of the cloned HIV-1 sequence insert at each nucleotide position. This was done by aligning each raw or Primer ID consensus sequence to the known sequence, determining if the call was correct, and totaling the number of erroneous nucleotide calls for all sequences for each nucleotide position. The percent of erroneous calls (errors/total nucleotides sequenced x 100) at each nucleotide position sequenced was plotted for both raw and Primer ID consensus sequences. Erroneous calls were further categorized as: (1) *substitutions*, where a nucleotide other than the known nucleotide was called; (2) *insertions*, where an extra nucleotide was called in that position; (3) *deletions*, where the nucleotide was represented as a gap in that position; or (4) *other errors*, such as an ambiguous base call.

For error rates, erroneous nucleotide calls were summed across all 317 interrogated nucleotide positions spanning HXB2 nucleotides 2648-2964, both for all copy number dilutions and separately for each copy number dilution (barcode). The total number of errors was divided by the total number of nucleotides read and multiplied by 10,000 to determine the error rate per 10,000 nucleotides. Error rates were calculated

for strata defined by sequencing platform (454 Junior or MiSeq), analysis method (raw or Primer ID consensus sequences), and by sequence location (homopolymeric or heteropolymeric regions). Standard errors and 95% confidence intervals (CI) for error rates were calculated using the Poisson distribution. Clustered sandwich estimators were used to calculate standard errors and 95% CI for error rates calculated across all copy number dilutions [351].

Sampling Minority Variants: The ability to detect minority variants depends on the number of unique Primer ID consensus sequences within a sample (Figure 3.2). Assuming an error rate of 4/10,000 nucleotides and a 400 nucleotide sequence, 59 Primer ID consensus sequences are necessary to detect a variant present in 5% of the population with 95%power in a hypothesis test with a 5% type I error probability. At 95% power, only 29 Primer ID consensus sequences would be needed to detect a variant present in 10% of the sample, while 298 Primer ID consensus sequences would be needed to detect a 1% variant.

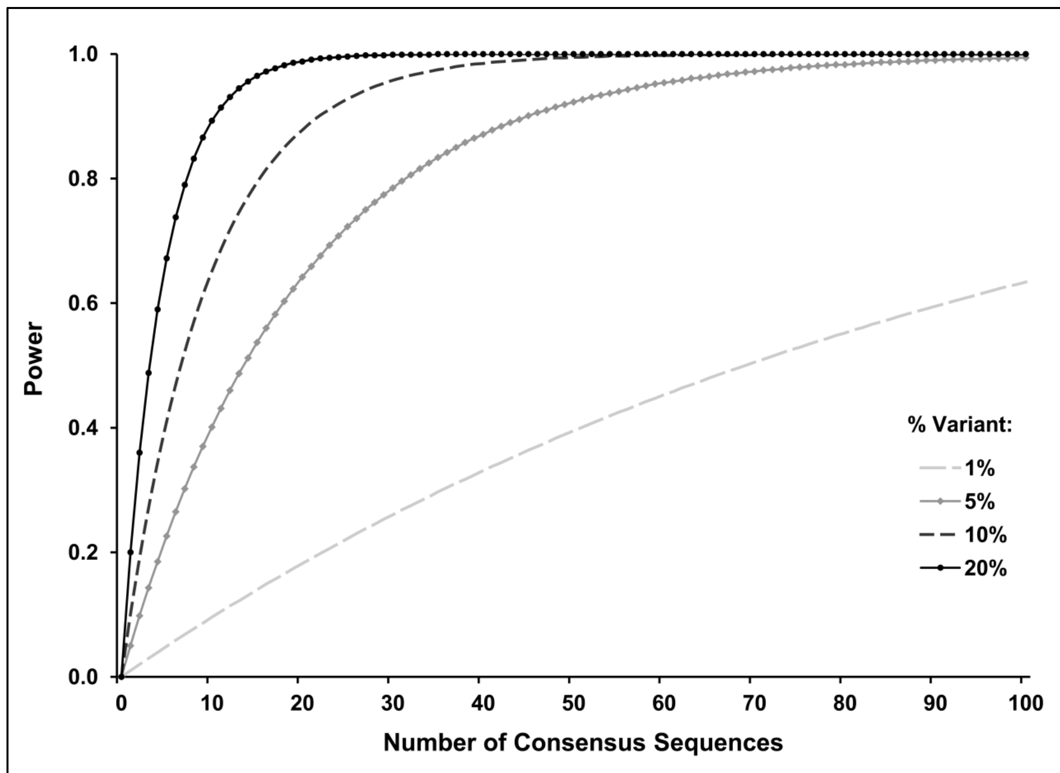

Figure 3.2. Sampling depth and power to detect minority HIV-1 sequence variants. Assume 4 errors per 10,000 nucleotides, a 400 nucleotide sequence, and α=0.05.

3.6.2 SPECIFIC AIM 2A/B

(A) Using the 454 FLX platform to sequence HIV-1 RT codons 34-138 and 149-236, estimate the prevalence and relative abundance of RT inhibitor resistance among 184 therapy-naïve patients with concurrent bulk sequencing results. (B) Compare estimates obtained using the 454 platform to estimates obtained using bulk sequencing.

Study Population: The study population for Aim 2A and 2B consisted of all participants enrolled in the UCHCC that met the inclusion criteria outlined below (Figure 3.1). All HIV-infected persons at least 18 years of age and providing written, informed consent for inclusion in the UCHCC were eligible if: (1) they initiated combination antiretroviral therapy after December 31, 1999; (2) with either ≥3 NRTI or ≥2 NRTI+NNRTI; (3) they had ≥1 HIV-1 RNA level measurement recorded in the database prior to therapy initiation; and (4) they had available frozen plasma prior to therapy

initiation. 331 patients met criteria (1) through (3), while 184 of these patients had available pre-therapy samples (4).

To examine exchangeability between populations with and without available samples with respect to pre-existing drug resistance, distributions of key variables were compared:

1. ***Age at therapy initiation***
2. ***Gender***
3. ***Race:*** black, white, or other race.
4. ***HIV risk category:*** separate binary variables men who have sex with men (MSM), injection drug use (IDU), and heterosexual contact.
5. ***Year of therapy initiation:*** categorized as 1999-2001, 2002-2004, 2005-2007, and >2007.
6. ***Last pre-therapy CD4+ T cell count:*** continuous or categorized using the CDC laboratory definition of AIDS (in cells/μL) [32]: <200; and ≥200.
7. ***Last pre-therapy HIV-1 RNA level:*** $\log_{10}$-tranformed and treated as continuous.

Distributions of continuous variables were first checked within all eligible patients (N=331) by inspecting boxplots and histograms, and non-normally distributed variables were transformed as necessary. Next, continuous and categorical variables were examined within patients with an available sample (N=184) and those without an available sample (N=147) and their differences compared.

Aim 2A Analysis: The main objective of Aim 2A was to estimate the amount of pre-existing resistance within a population of mostly chronically-infected, therapy-naïve patients using highly-sensitive UDS to measure resistance mutations. Resistance mutations were defined using the 2009 list of surveillance drug resistance mutations

(SDRM) (Appendix 2.2) since detection of these mutations in patients before starting therapy is more likely to represent transmission of drug resistance rather than naturally occurring polymorphisms [188, 189]. However, given high error rates associated with homopolymeric positions, mutations at HIV-1 RT codons 65, 67, 74, 100, 101, 103, 115, 116, and 219 were excluded from overall estimates of resistance prevalence.

Deep Sequencing Data: Two amplicons per patient were submitted for sequencing using the Roche 454 FLX deep sequencing platform: one amplicon spanned HXB2 nucleotides 2620-2992, and one spanned HXB2 nucleotides 2965-3284. However, the HIV-1 specific primer sequence was excluded from analyses since these primer sequences are theoretically identical across all sequences and not biologically informative. Therefore, the sequence regions used to estimate resistance in this population spanned HXB2 nucleotides 2648-2964 and 2993-3257, which cover HIV-1 RT codons 34-138 and 149-236, inclusive.

Not all patient amplicons were successfully sequenced using the 454 platform. For 7 (4%) and 3 (2%) subjects, only HIV-1 RT fragment 1 (RT codons 34-138) or 2 (RT codons 199-236) sequences were available, while for 2 (1%) patients, no sequences were obtained. For those missing 454 sequence information, it was assumed that they had no resistance within the missing RT region. Therefore, prevalence estimates used the full 184 patients as the denominator. Population prevalence was estimated overall (any RT inhibitor resistance), and by antiretroviral class (NRTI and NNRTI resistance). For population prevalence estimates, standard errors and 95% CI were calculated using the binomial distribution.

Deep sequencing resistance data for each patient consisted of the following: (1) sample draw date; (2) sample HIV-1 RNA level; (3) number of Primer ID consensus sequences for each HIV-1 region sequenced; (4) presence or absence of each SDRM

mutation among all their Primer ID consensus sequences; (5) number of Primer ID consensus sequences with each SDRM mutation; and (6) proportion of Primer ID consensus sequences with each SDRM mutation (relative abundance). In addition, class-wide and overall resistance variables were estimated per patient as follows: (1) presence or absence of any RT inhibitor, NRTI, and NNRTI resistance mutation; (2) the total number of unique RT inhibitor, NRTI, and NNRTI resistance mutations detected; (3) for each amplicon sequenced and for both amplicons together, the minimum and maximum total occurrences of any RT inhibitor, NRTI, and NNRTI resistance mutation among Primer ID consensus sequences; and (4) for each amplicon sequenced and for both amplicons together, the minimum and maximum proportion of Primer ID consensus sequences with any RT inhibitor, NRTI, and NNRTI resistance mutation.

In light of error rates observed in the control experiments from Aim 1, a stricter definition of deep sequencing detected resistance was implemented for a sensitivity analysis. For this analysis, all single occurrences of resistance mutations were excluded as errors. That is, within each patient sample, a resistance mutation had to occur on at least two Primer ID consensus sequences to be counted as genuine. This approach was probably overly conservative, however. For the upstream reverse transcriptase fragment, given 12,971 consensus sequences generated across 181 patients and assuming that errors follow the Poisson distribution, a total of 493 singly occurring mutations were expected; 1,887 were observed (observed/expected: 3.8). For the downstream fragment, 379 singly occurring mutations were expected given 10,122 consensus sequences across 177 patients, while 1,828 were observed (observed/expected: 4.8).

Analysis Aim 2B: Deep sequencing and standard bulk sequencing analysis resistance results were compared. Since standard bulk sequencing cannot reliably

detect minority variants that comprise less than 20% of the total mixture [146], all SDRM detected by bulk sequencing were expected to be detected using the 454 deep sequencing platform. Comparisons between ultra deep and bulk sequencing detected resistance were repeated for each definition of ultra deep sequencing detected resistance.

Bulk Sequencing Data: Bulk sequencing resistance data for each patient consisted of the following: (1) date of sample draw; (2) sample HIV-1 RNA level; (3) an indicator variable for whether or not bulk sequencing was done in-house; (4) laboratory where sequencing was performed; (5) presence or absence of each SDRM mutation; (6) presence or absence of any RT inhibitor, NRTI, and NNRTI mutations; and (7) number of unique RT inhibitor, NRTI, and NNRTI mutations.

Agreement between bulk sequence analysis and deep sequencing was assessed for each SDRM and for each antiretroviral class by calculating Kappa statistics (-1 to 1). Median relative abundance of SDRM was compared by whether or not the mutation was detected using standard bulk sequencing.

Statistical Precision: Using the logit transformation of proportions and their standard errors to calculate confidence intervals, precision around prevalence estimates increases with lower prevalence and increasing sample size (Figure 3.3). The logit transformation was used to avoid illogical confidence intervals that fall outside the 0 and 1 range. Assuming a sample size of 184 patients, prevalence estimates as low as 2% may be statistically distinguished from zero (α=0.05).

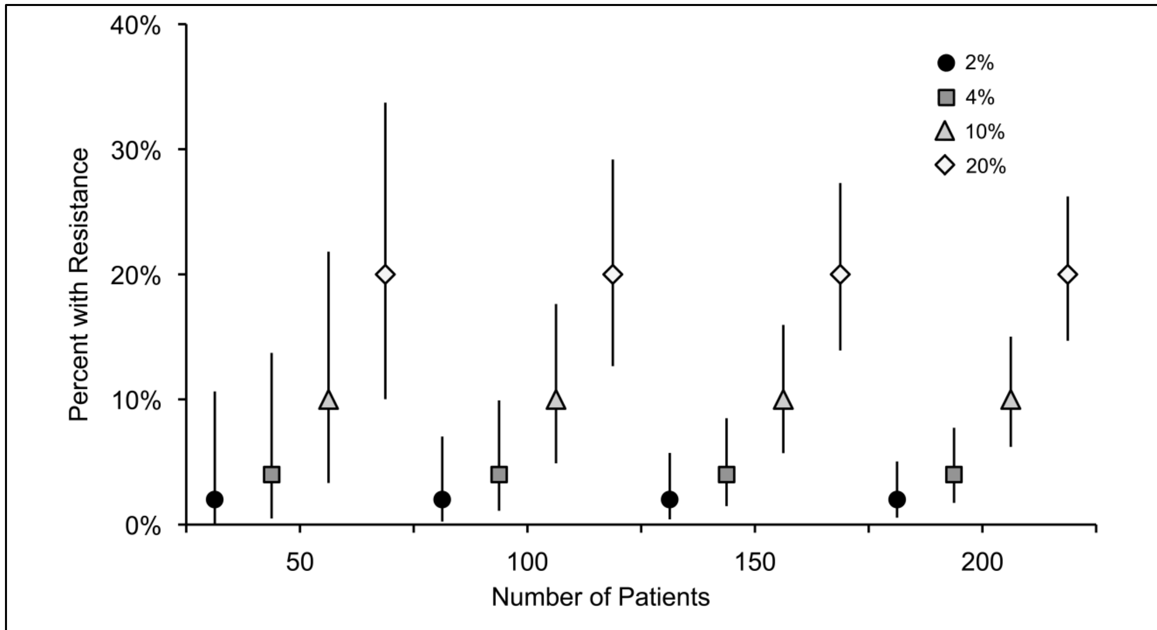

Figure 3.3. Prevalence and 95% confidence intervals by sample size, assuming 2%, 4%, 10%, or 20% of the study population has resistance. Confidence intervals were calculated using the logit transformation.

3.6.3 SPECIFIC AIM 2C

(C) Among a subset of 19 patients with the greatest depth of sampling of viral templates, compare mutations detected within HIV-1 RT codons 34-73 and 111-138 between the Illumina MiSeq and Roche 454 platforms.

Study Population: Initially, 22 patients were chosen from the full population of 184 patients for sequencing using the Illumina MiSeq platform, which is not susceptible to homopolymer-associated error. This subset was chosen since they had the greatest number of Primer ID consensus sequences generated using the 454 platform, and their barcode sequences were at least two nucleotides different from each other, making it less likely sequences would be misclassified. However, adequate amplified DNA from the previous 454 deep sequencing experiment was available for only 19 patients, who had a median 203 Primer ID consensus sequences (IQR: 168-243) generated from a median 3,378 raw 454 sequences (IQR: 2,494-3,796). Amplified DNA spanning HXB2 nucleotides 2620-2992 from the second round of PCR was pooled in equal amounts

across 19 patients, gel purified, and submitted for deep sequencing using the 150 bp paired-end sequencing protocol on the Illumina MiSeq sequencing platform. Using this protocol, sequences spanning HXB2 nucleotides 2648-2840 and 2782-2964 were available for comparison.

Analysis: The objective of Aim 2C was to compare any detected mutations between two different ultra deep sequencing methods: Illumina MiSeq and Roche 454 FLX platforms. While Illumina is not susceptible to homopolymer-associated miscalls, it has a shorter read length limitation, and it is associated with its own set of platform specific errors, discussed in Section 2.9.3. The majority of singly occurring mutations within the 454 Primer ID consensus sequences were expected to fall within homopolymeric regions. In contrast, for Illumina MiSeq, the quality of raw sequences was expected to decline over the sequencing run, so more single occurrences of mutations were expected towards the end of each paired-end read. In this case, more single occurrences of mutations would be expected towards RT codon 73 (downstream) for one paired end, while for the other paired end, more single occurrences would be expected towards RT codon 111 (upstream).

Deep Sequencing Data: The upstream HIV-1 reverse transcriptase amplicon for each of 19 patients (spanning HXB2 nucleotides 2620-2992) was submitted for sequencing using the Illumina MiSeq deep sequencing platform with the 150 bp paired-end sequencing protocol. Thus, the biologically relevant sequence did not completely cover the entire amplicon (317 bp). Therefore, the sequence regions used for comparison to 454 deep sequencing spanned HXB2 nucleotides 2648-2770 and 2878-2964, which covered HIV-1 RT codons 34-73 and 111-138, inclusive.

Deep sequencing data for each patient consisted of the following: (1) sample draw date; (2) sample HIV-1 RNA level; (3) number of Primer ID consensus sequences

for each platform; (4) presence or absence of each mutation among Primer ID consensus sequences for each platform; (5) number of Primer ID consensus sequences with each mutation for each platform; and (6) proportion of Primer ID consensus sequences with each mutation (relative abundance) for each platform. Sequences were first examined at the amino acid level, and when differences between platforms were found, at the nucleotide level.

For each patient, the number of Primer ID consensus sequences generated was compared by sequencing platform. Even though the Illumina MiSeq capacity is much greater than the 454 sequencing platform (10-fold increase in raw sequences), a comparable number of Primer ID consensus sequences was expected per patient since the same PCR amplicons were sequenced. After comparing these values, amino acid mutations were compared across platforms both by individual patient, and overall across all patients. Mutations were categorized as detected by MiSeq alone, by 454 FLX alone, and by both platforms. Median relative abundances of mutations detected solely by Illumina MiSeq, solely by 454 FLX, and by both sequencing platforms were compared. For each mutation detected, the nucleotide sequence was examined to determine its distance from a homopolymeric tract. Also, their location along HXB2 was plotted by the number of times the mutation was detected.

3.6.4 SPECIFIC AIM 3

Among study participants with RT inhibitor resistance measured using both bulk and deep sequencing, estimate the effect of pre-therapy resistance to one or more antiretrovirals within their first regimen on time-to-first virologic failure (VF).

Study Population: The effect of pre-therapy resistance on treatment response (Aim 3) was estimated among a subset of Aim 2A/B's population that had at least one HIV-1 RNA level measurement after beginning HAART. Of 184 patients with pre-therapy samples available for resistance testing (Aim 2A/B), 153 had at least one HIV-1 RNA level measurement following HAART initiation.

Statistical Model: Virologic response to HAART is indicated by a single biomarker that is repeatedly measured both prior to and after HAART initiation: HIV-1 RNA level in the plasma, which is expected to decline to below detection (<50 HIV-1 RNA copies/mL) with successful antiretroviral treatment. Treatment failure is not defined by guidelines using the slope of HIV-1 RNA level decay within a defined time period; instead individuals experience treatment failure when plasma HIV-1 RNA levels rise above or remain above 200 copies/mL 2-8 weeks following initial treatment or regimen switch [6]. The timing of HIV-1 RNA replication suppression is important, however, since individuals who more rapidly reach these endpoints experience better outcomes in terms of decreased evolution of drug resistance [167], enhanced immune reconstitution [84], and decreased mortality [175]. Therefore, the definition of VF used here included an HIV-1 RNA slope change from baseline so that patients who failed to achieve at least a 0.5 $\log_{10}$ drop in HIV-1 RNA copies/mL from baseline within the first 6 months of therapy were also considered to have experienced VF, rather than just patients who experienced a HIV-1 RNA level above detection (in this case >400 co/mL) after 6 months of therapy. Since the outcome (VF) is not only the occurrence of virologic response but also the

timing of response, represented by event time ***T***, a Cox proportional hazards model of the general form was used [352]:

$$\lambda_T[t \mid X(0), L(t)] = \lambda_0(t) e^{\beta_1 X(0) + \beta_2 L(t)}$$

where $\boldsymbol{\lambda_0(t)}$ is the unspecified baseline hazard function for first VF for individuals without evidence of baseline resistance to their initial regimen and with all covariates in vector ***L(t)*** set to 0. $\boldsymbol{\beta_2}$ is a vector of unknown coefficients for time-fixed (measured at baseline, ***t=0***) and time-varying covariates ***L(t)***, while $\boldsymbol{\beta_1}$ represents the unknown coefficient for the main exposure (pre-existing resistance to their initial regimen). Cox's partial likelihood method for estimating the conditional hazard of event ***T*** at time ***t*** may be extended so that the effects of variables may vary over time (relaxing the proportional hazard assumption) and so that individuals may enter and leave the risk set during the study period. However, since partial likelihood methods depend on the ordering of events, rather than the timing of events, ties in event times can be problematic. For our purposes, we relied on the exact methods for partial likelihood estimation of unknown beta coefficients proposed by Kalbfleisch and Prentice in 1980 [353], which assume that there are no real tied event times and that the true ordering of events is unknown because of imprecise measurement of time. If exact computations became too time-consuming, approximations proposed by Efron were used [354], which are usually closer to exact estimations than other approximate methods [355].

Exposure Variables: To construct clinically relevant exposure variables for pre-therapy resistance, patients' pre-existing resistance mutations were interpreted in the context of the combination of RT inhibitors they received over follow-up rather than examining the effect of any pre-existing resistance regardless of therapy received. To interpret individual patient genotypes, we relied on the Stanford HIV Drug Resistance Database (HIVdb) algorithm shown in Appendix 2.3 [171], which is also available online

at http: //hivdb.stanford.edu/. For bulk sequencing results, each patient's amino acid sequence was uploaded to the Stanford HIVdb, which returned a susceptibility score for each of the following RT inhibitors: 3TC, FTC, ABC, AZT, d4T, ddI, TDF, EFV, ETR, NVP, and RPV. These scores were discrete values in intervals of 5, ranging from 0-5 (susceptible), 10 (potential low-level resistance), 15-40 (intermediate resistance), and 45 and above (high-level resistance). Individual antiretroviral scores were compared to each patient's initial regimen, and patients receiving ≥1 antiretroviral (ARV) with a HIVdb score over 10 (≥15) were considered to have received a partially active initial regimen. For UDS genotypes, the above procedure was applied to patient Primer ID consensus sequences; however, Primer ID amino acid consensus sequences were interpreted within the context of homopolymer-associated errors.

Susceptibility to initial regimen was categorized for each patient as follows: (1) <3 active ARV vs. ≥3 active ARV versus using resistance detected either bulk sequencing or UDS; (2) by relative abundance (minority or dominant vs. ≥3 active ARV) using resistance detected using UDS; and (3) by absolute copy number of resistant sequences, calculated using the proportion of resistant consensus sequences and sample viral load. For patients with multiple RT inhibitor resistance mutations, the highest relative abundance among these was used to calculate the number of resistant copies. Finally, resistance was categorized by sequencing detection method (<3 active ARV by UDS only, by <3 active ARV bulk sequencing and UDS vs. ≥3 active ARV). This last variable was constructed to provide some insight into the added value of UDS compared to the standard method of resistance measurement. Each variable was considered overall (<3 active RT inhibitor), and by ARV class (<3 active NRTI or any <3 active NNRTI).

Homopolymeric Regions: Again, deep sequencing resistance genotypes were expected to be susceptible to homopolymer-associated errors, given that deep sequencing was performed using 454 FLX platform. To account for homopolymer-associated errors, homopolymeric sequencing regions were defined as sequences made up of 3 or more identical nucleotides in a row plus the two flanking nucleotides, which is more conservative than the definition given for Aim 1. Applying this definition to the 2004 HIV-1 Subtype B consensus sequence, 140 (44%) of 317 and 94 (35%) of 265 nucleotide positions were within homopolymeric influence within the upstream and downstream sequences. From the HIVdb drug resistance mutations listed in Appendix 2.3, amino acid positions at HIV-1 RT codons 40, 41, 44, 65, 66, 67, 69, 70 71, 74, 77, 98, 100, 101, 103, 115, 116, 118, 151, 188, 215, 219, and 227 were considered susceptible to homopolymeric error. Mutations on these homopolymer-susceptible codons had to occur on at least 2 Primer ID consensus sequences within a patient sample to be considered genuine. In a sensitivity analysis, this more conservative requirement was extended to all resistance positions.

Outcome Variable: Patients were followed from the date of HAART initiation (origin) until the earliest HIV-1 RNA measurement at or before the first of the following: (1) switch to a regimen containing a non-RT inhibitor ARV; (2) discontinuation of HAART for 2 weeks or more (4 weeks in a sensitivity analysis); (3) the end of the follow-up period, August 17, 2012; or (4) virologic failure (VF), where VF was defined as failure to achieve at least 0.5 $\log_{10}$ HIV-1 RNA copies/mL decrease from baseline within the first 6 months of HAART, or at 6 months or later, an HIV-1 RNA level of 400 copies/mL. The threshold of 400 copies/mL was chosen since this value was the highest limit of detection within the database.

Confounders: Confounders of the association between pre-therapy resistance and time-to-first virologic failure were first identified using a directed acyclic graph (DAG), shown in Figure 3.4. The exposure of interest is pre-therapy resistance to one or more antiretrovirals within the initial HAART regimen, represented by $X(0)$ in the DAG. $X(0)$ was technically measured using the sample most proximal but before HAART initiation, however, since the origin of time is HAART initiation and not the time of exposure measurement, the effect of $X(0)$ was allowed to vary with $S(t)$ to account for this lag. The outcome, $\lambda_T(t)$, is the hazard of virologic failure, while T is the random variable representing individuals' failure/censoring times. All repeated measures, $L_n(t)$, and unmeasured time-varying adherence and treatment indication, $U(t)$, where t indexes the timing of each measurement, lie on the causal path from $X(0)$ to $\lambda_T(t)$. Therefore, effect-measure estimates were not conditioned on these variables. However, variables included in sets Z (race/ethnicity, gender, and HIV risk group, and age at HAART initiation), and the time from resistance measurement until HAART initiation, $S(t)$, lie on non-causal, unblocked backdoor paths from $X(0)$ to $\lambda_T(t)$, and are therefore should be adjusted for in the analysis. Adequate control may be achieved by only adjusting for variables in set Z, since the only pathway between $S(t)$ and the outcome would be blocked by adjusting for Z.

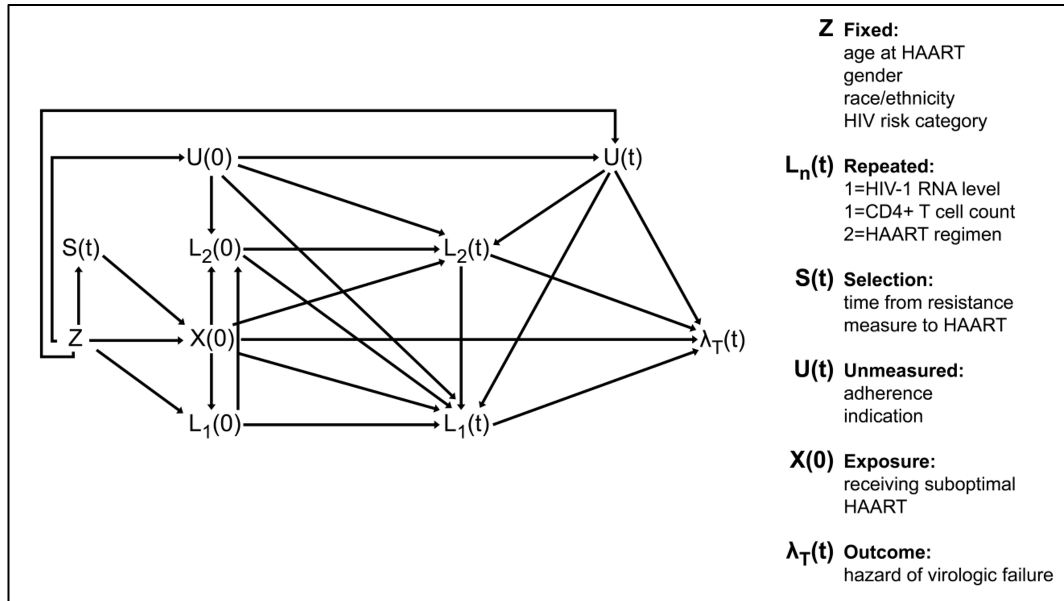

Figure 3.4. Directed acyclic graph of the causal association between pre-therapy resistance and time-to-virologic failure. Confounding through indication includes treatment decisions made by care providers.

Confounder variables were included in the model as follows:

(1) Age (Z): age at therapy initition was modeled using restricted cubic splines with knots at the quartiles of the entire population distribution, allowing the effect of age to vary in a non-linear fashion.

(2) Gender (Z): female versus male.

(3) Race (Z): white versus non-white race.

(4) MSM (Z): MSM versus non-MSM. IDU was not included since only 15 patients fit into this category and none were exposed.

(5) Lag Time [S(t)]: Days between the draw date of the sample used for UDS and therapy initiation was modeled as a simple continuous variable.

For each confounding variable and for each resistance exposure variable, a cumulative incidence curve was created to examine the relationship between the confounder and cumulative incidence of virologic failure [356]. Curves were compared between categories of confounding variables using the log-rank test. For age and time

between UDS sample draw and HAART initiation, categories defined using quartile cut-points were used to examine cumulative incidence curves.

Proportional Hazards Assumption: Proportional hazards models, unadjusted and adjusted and stratified by months from UDS sample draw to therapy initiation, were fit for each resistance variable. Hazard ratios (HR) and 95% confidence intervals (CI) compared the hazard rate of VF among patients with resistance to their initial regimen, measured at the time of therapy initiation, to those initiating with all active antiretrovirals. The proportional hazards assumption was evaluated by inspection of log[-log(survival)] curves for confounder variables and for each exposure variable. If graphs of the survival function versus the log of survival time revealed non-parallel lines for a variable, an interaction with time since HAART initiation was included in the model. Using this criterion, the effects of MSM and race were allowed to vary over time since HAART initiation.

Non-Informative Censoring: For sensitivity analyses to quantify the influence of potentially informative censoring, the regression analysis was repeated under two extreme operating assumptions [357]: (1) Censored observations were assumed to experience VF immediately after censoring. (2) Individuals who were censored had event times greater than any individuals with observed event times. For the first analysis, all patients were assumed to have experienced VF on the day after they were censored. For the second assumption, the event time for censored observations was changed to the maximum event time plus one day, August 18, 2012. Analyses, both adjusted and unadjusted, were repeated for each assumption and effect-measure estimates were compared between sets of three models to examine the influence of these assumptions on effect estimates.

Statistical Power: Assuming 10% of the 184 patients received less than fully active HAART, 25% of the patients without any resistance experience virologic failure within one-year, and 30% of patients are censored during follow-up, this study would have 80% power to detect an HR of 2.1 (α=0.05, two-sided log-rank test). HR estimates are from models containing the only the main exposure, or for models where there is no correlation between additional confounding variables and the main exposure. Power curves were generated assuming a range of individuals experience the event (Figure 3.5). Here, statistical power declines as the proportion who experience the event declines as well as with the HR for the main effect.

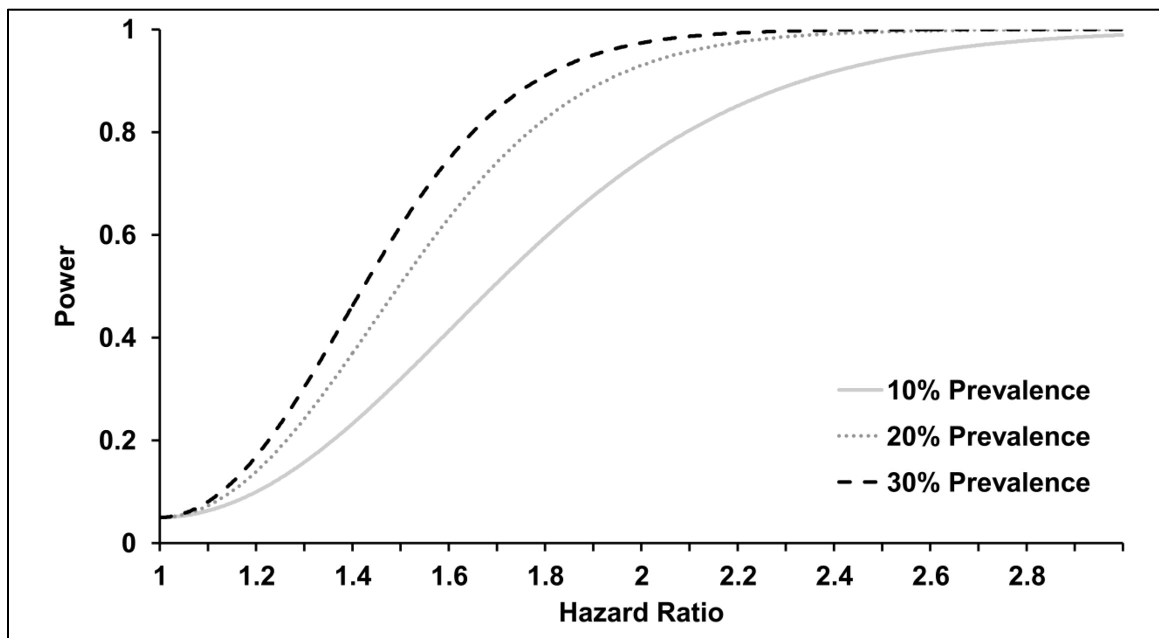

Figure 3.5. Statistical power to detect various hazard ratios, assuming 30% non-informative censoring, virologic failure occurs among 25% of unexposed individuals within one year of HAART, and 10%, 20% or 30% of the study population has evidence of pre-existing drug resistance, two-sided test, *α=0.05.* Power curves are estimated for regression models containing the main exposure only. HAART=highly active antiretroviral therapy.

3.7 Limitations

The limitations of the proposed study affect each of the aims, and may be divided broadly by their relationship to validity and precision. Validity-related limitations include restricted generalizability, information bias, unmeasured confounding, and selection

bias. Precision-related limitations are associated with the fixed sample sizes, both in terms of the sampled patient population and the sampling depth achieved for minority viral variants. These limitations are outlined below:

Generalizability: The results of our analysis may not be generalized to populations outside the Southeastern US, especially to areas where resistance testing is not routinely offered to all therapy-naïve individuals as they enter care. However, the results could be extrapolated with caution to other populations where the HIV epidemic includes more rural residents, among minority populations and driven by sexual transmission, and where the majority of individuals are diagnosed with HIV later in the course of infection. In addition, this analysis is restricted to RT inhibitor resistance, which limits generalizability to these two classes of antiretrovirals only. Since these antiretrovirals continue to make up the majority of first line regimens worldwide, this limitation is not particularly restrictive at this time.

Information bias: Information bias was primarily related to exposure measurement by UDS in terms of accounting for additional sources of sequence error and achieving adequate sampling of the viral population. On one level, 454 deep pyrosequencing is superior to standard bulk sequencing given increased sensitivity for minority variants, quantification of variants in the viral population, and the possibility of linkage analysis; however, resistance data must be interpreted carefully to properly account for any additional errors introduced by this platform that may increase false positives (i.e., reduce specificity) and that may increase false negatives (i.e., reduce sensitivity). To limit the impact of this first limitation, the Primer ID method was used to more accurately estimate the viral population, rather than the population that best amplified, and resistance among the clinical population was interpreted within the

context of error rates from control experiments by imposing stricter definitions to account for errors associated with homopolymeric sequence regions.

With respect to sampling depth, Primer ID enumerates the viral population within the infected patient, but the ability to detect minority variants is sharply limited by the number of viral templates sampled. To address this problem, several laboratory procedures were adjusted to maximize the number of templates sampled per patient. For example, virus was concentrated prior to RNA extraction if sample viral load was $<4.5 \log_{10}$ HIV-1 RNA copies/mL. However, preliminary experiments reveal that the number of viral genomes that are actually sampled is a small fraction of the total number that are available, assuming accurate viral load measurement.

Unmeasured Confounding: There are at least two sources of unmeasured confounding in this study that may not be adequately controlled with measured confounders in our analysis for Aim 2: (1) adherence to HAART; (2) confounding by indication, whereby treatments are selected based on provider-patient decisions (disease severity, physician experience, underlying health of patient, etc.). On the DAG in Figure 3.4, the path from these unmeasured confounders adherence and indication are blocked by adjustment for measured confounders in the set ***Z***, including age at HAART initiation, gender, race/ethnicity, and HIV risk factors. While we have no surrogate information about unmeasured adherence, we are able to examine the effect of one specific type of confounding by indication. Some patients in this dataset may have been offered alternative HAART regimens based on pre-therapy bulk sequencing (N=37). None of the 37 patients had reduced susceptibility to any of their prescribed regimens as predicted using bulk sequence analysis, but a few did have predominant resistance mutations that were detected by standard sequence analysis. This phenomenon itself may introduce bias if patients with drug resistance at baseline also

tended to have minority drug resistant variants that would not be detected by standard bulk sequence analysis.

Selection Bias: Selection bias may be an important source of error for analyses involving clinical samples, since almost half of samples from eligible patients were not available for resistance testing. Patients with available samples tended to have a longer amount of follow-up time from entry to care and HAART initiation, and they had more samples drawn prior to HAART. This may lead to selection bias due to underrepresentation of fast progressors, which would artificially lengthen virologic failure times. In addition, the lack of fast progressors could inflate the estimate the prevalence of pre-therapy resistance, since individuals with infected with resistant virus may experience some initial benefit in terms of higher CD4 cell count [304], which could bias the hazard ratio estimates upward. However, if fast progressors were more likely to carry resistant virus, particularly multi-drug resistant virus, the effect-measure estimates for Aim 2 could be biased downward.

Statistical Precision and Power: The precision of our measures of disease frequency and effect-measure estimates are limited by the fixed sample size for each study population, in terms of the population of patients. To address the problem with patient sample size, only retrospectively obtained samples were tested for UDS, and many of these were not available. A concerted effort was made to obtain the maximum number of clinical samples; however, the size of this population is relatively large considering the increased sensitivity and coverage of UDS.

CHAPTER 4 PRIMER ID CORRECTS NEXT-GENERATION SEQUENCING PLATFORMS AND STILL REVEALS PRE-EXISTING DRUG RESISTANCE MUTATIONS IN THE HIV-1 REVERSE TRANSCRIPTASE CODING DOMAIN

4.1 INTRODUCTION

Combination antiretroviral therapy continues to improve patient outcomes as better treatment options are developed [1, 358, 359]. Advances may be offset among participants failing multiple regimens by development of resistance and cross-resistance [116, 144], which also may be transmitted to susceptible partners [360, 361]. Since transmitted drug resistance may compromise patient response to first-line combination therapy [303, 362, 363], genotypic resistance testing is routinely recommended before therapy initiation [6]. The utility of pre-therapy testing may be limited by minority HIV-1 variants, present in <20% of the viral population, that are not reliably detected by standard sequencing [146], and that may jeopardize virologic response [346, 364-367].

The prevalence of minority pre-therapy drug resistance varies, with estimates based on highly-sensitive research assays often double those reported using standard sequence analysis [336, 346, 365-368]. Some variation may be related to methods for measuring low abundance resistance. For example, allele-specific PCR detects mutations that make up ≥0.01% of the viral population [366]; however, prevalence estimates are based on a few, pre-determined mutations, and estimates of resistance in an individual may be biased by differential amplification. Alternatively, single genome sequencing allows interrogation of entire viral genomes diluted to a single viral sequence, bypassing some amplification errors, but this labor-intensive method generally achieves low sensitivity due to limited sampling depth [146]. In contrast, ultra deep sequencing involves massively-parallel sequencing of

amplified viral sequences, producing upwards of 10^6 sequences and reaching apparent sampling depths of 1% or less [22]; however this method frequently generates errors during amplification and sequencing, making it difficult to distinguish true minority variants from sequencing errors [24]. In addition, resequencing of a smaller number of viral genomes after PCR amplification (PCR resampling) gives over-estimates of the true sampling depth.

To address errors associated with deep sequencing, others have established threshold cut-offs based on estimated error rates from known sequences [23]. Cut-offs do not account for errors that may be introduced during the PCR step, such as biased amplification and nucleotide misincorporation [19, 20], nor do they address PCR resampling. Here, an alternative strategy engineered into the cDNA synthesis step (Primer ID) was used to circumvent the need for statistically-defined cut-offs [26] allowing: (1) definition of a background error rate for two deep sequencing platforms; (2) estimation of the prevalence of pre-existing reverse transcriptase inhibitor (RTI) resistance among chronically-infected patients; and (3) comparison of prevalence estimates from standard methods to those obtained by deep sequencing.

4.2 METHODS

4.2.1 STUDY POPULATION

Study participants were previously enrolled in the University of North Carolina Center for AIDS Research HIV Clinical Cohort Study (UCHCC) [339]. UCHCC is an ongoing, clinical cohort enrolling HIV-infected adults receiving care at UNC. UCHCC maintains an electronic database of patient information and houses a repository of plasma samples obtained during routine care. Patients were eligible for this study if they: (1) provided informed consent for inclusion in the UCHCC; (2) initiated therapy after December 31, 1999; (3) with two or more nucleoside/tide reverse transcriptase inhibitors (NRTI) plus one non-nucleoside reverse transcriptase inhibitor (NNRTI), or three or more NRTI; (4) had at least

one reported pre-therapy HIV-1 RNA level; and (5) had an archived pre-therapy plasma sample available for resistance testing. This study was reviewed and approved by the University of North Carolina Institutional Review Board.

4.2.2 HIV-1 Sequencing

For most participants (141/184), bulk sequencing analyses were obtained using commercial HIV-1 GenoSure (Plus) assays (LabCorp, Research Triangle Park, NC). If no bulk sequence analysis was available (43/184), we attempted in-house sequencing of HIV-1 reverse transcriptase (RT) codons 34-245 using the ABI Prism BigDye Version 1.1 Terminal Cycle Sequencing (Life Technologies, Carlsbad, CA). To check for evidence of cross-contamination, sequences were aligned by ClustalW version 2 [347] and inspected by constructing Neighbor-Joining phylogenetic trees evaluated with 1,000 bootstrap replicates [348].

Sample amplicon libraries were generated using previously described methods [26]. Samples with <4.5 log_{10} HIV-1 RNA copies/mL were centrifuged to concentrate the virus particles prior to RNA extraction (QIAamp viral RNA extraction kit, QIAGEN, Hilden, Germany). One-third of the RNA was used in separate cDNA synthesis reactions targeting two regions of HIV-1 RT, HXB2 nucleotides 2648-2964 and 2965-3257 (RT codons 34-139 and 139-245) [346], using the primers listed in Appendix 3.1. The cDNA primers included a barcode to allow pooling of samples during the deep sequencing step, and a randomized sequence tag of 8 nucleotides (Primer ID) to allow identification of each individual template in the subsequent sequence analysis (Figure 4.1A). Purified cDNA [26] was used for semi-nested PCR (Figure 4.1B) using Phusion Hot Start II High Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA); annealing temperatures were 67°C and 63°C for RT fragments 1 and 2, respectively (Appendix 3.1). Input cDNA was estimated based on the assumption that all RNA templates (500 copies) were copied into cDNA.

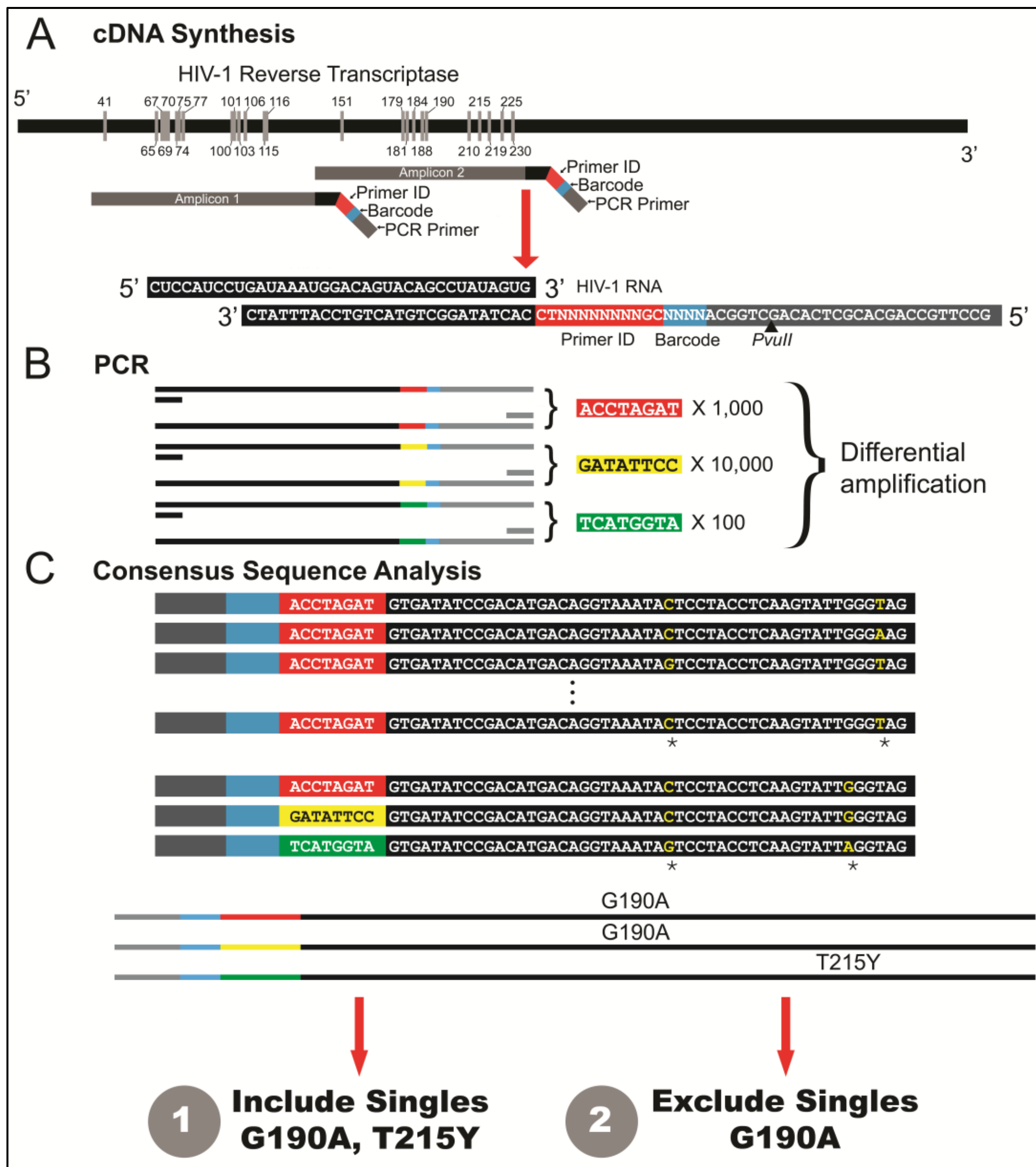

Figure 4.1. Primer ID to estimate the HIV-1 population. (A) During cDNA synthesis, a unique Primer ID sequence is incorporated into each viral genome along with a sample-specific barcode. (B) During PCR, differential amplification may occur so that the probability of amplification is not equally distributed across viral genomes. (C) Primer ID is applied to correct errors that accumulated over a deep sequencing run. Within each sample, a single majority-rules consensus sequence is generated from 3 or more sequences with the same Primer ID (red, green, yellow). Collectively, consensus sequences reflect the number of viral genomes rather than what best amplified. Method (2) is more conservative than (1) since single occurrences of resistance mutations on Primer ID consensus sequences are also excluded as error.

Amplified DNA was pooled in equimolar concentration, cleaved with *PvuII* to remove part of the PCR primer, and gel purified (QIAquick gel purification kit, QIAGEN). Pools were submitted for sequencing on the 454 GS FLX sequencing

platform with XLR80 Titanium reagents (Roche, Indianapolis, IN). To compare 454 and MiSeq nucleotide calls, 19 amplicons were also sequenced using the 150 bp paired-end sequencing protocol (HIV-1 RT codons 34-74 and 111-139, HXB2 nucleotides 2648-2770 and 2878-2964) on the Illumina MiSeq sequencing platform (San Diego, CA).

4.2.3 Plasmid Controls

The entirety of HIV-1 RT (HXB2 nucleotides 2550-3515), derived from a clinical sample, was cloned into vector pcDNA3.1 (Life Technologies). Plasmids were linearized with *BamHI*, purified using the Minelute PCR purification kit (QIAGEN), quantified by UV spectrophotometry using a Nanodrop 1000 (Thermo Fisher Scientific), and serially diluted to 3,000, 10,000, 30,000, and 300,000 copies. Plasmid DNA dilutions were denatured at 95°C for 5 minutes, cooled, and tagged with distinct cDNA primers during a single round of DNA synthesis with Platinum Taq (Life Technologies). Excess cDNA primers were removed [26], samples were amplified by nested PCR using a protocol identical to samples with primers listed in Appendix 3.1, pooled in equimolar concentration, and gel purified (QIAGEN). Pools were sequenced over HIV-1 RT codons 34-139 using the Roche 454 Junior (HXB2 nucleotides 2648-2964) and Illumina MiSeq sequencing platforms (HXB2 nucleotides 2648-2868 and 2782-2964).

4.2.4 Sequence Analysis

Deep sequencing data was processed using a custom pipeline of computer programs [26]. Briefly, sequence length distributions were inspected, short reads were discarded, and the remaining sequences were compared to HXB2 *pol* for orientation and location. Sequences with an invalid Primer ID or barcode were discarded, and filtered sequences were partitioned first by barcode (sample) and then Primer ID (viral template). Sequences with a Primer ID that occurred less than

three times within a sample were discarded (fewer than five times for MiSeq), and consensus sequences were generated from sequences with the same Primer ID. A larger number of identical Primer ID reads was used to build the consensus sequence for data from the MiSeq platform because of the larger number of reads available and with the goal of obtaining a more reliable consensus sequence. There was no evidence that fortuitous consensus sequences were created as the result of sequencing errors in the Primer ID.

Drug resistance was defined using an updated list of surveillance drug resistance mutations to exclude polymorphisms that may not contribute to resistance phenotype (Appendix 2.2) [189]. Deep sequencing detected surveillance drug resistance mutations were quantified using two conditions (Figure 4.1C): (1) barcode- and Primer ID-defined consensus sequences; and (2) barcode- and Primer ID-defined consensus sequences, but only including resistance mutations that occurred more than one time within a sample (i.e. associated with two or more different Primer ID consensus sequences). Relative abundance of individual resistance mutations per sample was calculated by dividing the number of sequences with resistance by the total number of consensus sequences obtained for the sample.

Deep sequencing detected resistance mutations were considered in the context of homopolymeric regions, which are error hotspots for 454 sequencing platforms [24, 369]. We defined homopolymeric regions as four or more consecutive, identical nucleotides plus the two flanking nucleotides. We used the 2004 HIV-1 subtype B consensus sequence to define homopolymer-associated positions for clinical subject samples (available from the Los Alamos HIV Database at http: //www.hiv.lanl.gov), and we defined homopolymeric regions directly for the subtype C

plasmid control. In the subtype C control sequence, 75 homopolymer-associated positions were identified within the sequence spanning HXB2 2648-2964. In the 2004 subtype B consensus sequence, 61 and 45 homopolymer-associated positions were identified within the analogous upstream sequence and the downstream sequence spanning HXB2 2993-3257. Since mutations on HIV-1 RT codons 65, 67, 74, 100, 101, 103, 115, 116, and 219 were within homopolymeric influence, they were excluded from prevalence estimates.

4.2.5 Statistical Analysis

Standard errors and 95% confidence intervals (CI) for plasmid control error rates (errors per 10,000 nucleotides) were calculated across all samples using clustered sandwich estimators [351] and the Poisson distribution. Standard errors and 95% CI for proportions were estimated using the binomial distribution. Sequencing depth, or the number of sequences required to observe x% viral variant with 95% confidence, was estimated using a power analysis (Figure 3.2). Distributions of categorical variables were compared using Pearson's χ^2 test, median values of continuously distributed variables were compared using the Kruskal-Wallis test. Statistical analyses were conducted in SAS version 9.3 (SAS Institute, Cary, NC).

4.3 Results

4.3.1 Quantifying Deep Sequencing Error

The goal of this study was to examine low level resistance to RT inhibitors encoded within the RT coding domain. Control amplicons were designed to include several important resistance positions: RT codons 34-139 or HXB2 nucleotides 2648-2964. First, we established a residual error rate for both the 454 and MiSeq platforms using plasmid controls to evaluate our ability to interpret rare variants using either the raw sequences or sequences corrected by Primer ID, which was used as

the primer in the first round of DNA synthesis. We used Taq DNA polymerase rather than RT in the first round of synthesis because of low template utilization by RT when starting with a DNA template; also, we used a DNA template rather than an RNA template to avoid misincorporation during the synthesis of RNA in vitro. A total of 112,108 reads of the 317 nucleotide amplicon obtained using the 454 platform were collapsed into 2,893 Primer ID consensus sequences. The overall error rate using raw sequences was 71/10,000 nucleotides (95% CI: 70-72), which was reduced to 2.6/10,000 nucleotides (95% CI: 2.2-3.2) using the Primer ID/consensus sequence approach (Appendix 4.1). Over 75 homopolymeric positions, 6.0 (95%CI: 4.8-7.4) miscalls were observed every 10,000 nucleotides using Primer ID; excluding homopolymeric regions reduced error to 1.6/10,000 (95% CI: 1.3-2.0) nucleotides. Errors were substitutions (76%), deletions (22%), or insertions (1.7%). Error frequency is compared for each position queried in Figure 4.2A.

We were especially interested in the potential impact of homopolymeric error in the region of the K65 codon, the position of an important resistance mutation for tenofovir. In most subtype B isolates the lysine codon at this position is AAA and part of a longer homopolymeric region. However, some subtype B and most subtype C isolates have the AAG codon embedded in this longer homopolymer, and misplacement of the G can create the appearance of the AGA codon at this position which would be interpreted as a resistance mutation (arginine). In the control plasmid, which had a K65 AAG codon, only 59% of raw sequences had the correct sequence and 38% of the sequences were “ATA AAA A-G AAA GAC.” This was caused by the under-call of an A in the homopolymeric region in front of the AAG codon, thus shifting the G one position to the left and creating an AGA codon. Since it is not possible to know which A of the homopolymeric tract was undercounted one

cannot interpret the presence of the AGA codon. A more difficult situation would be when there is an under-call in the upstream portion of the homopolymeric tract and an over-call in the downstream portion, which would shift the G and leave the reading frame intact, thus obscuring the effect of the homopolymeric errors and creating an erroneous call for a resistance mutation. This was a very rare event that occurred in 12/112,108 raw reads.

Alternatively, deep sequencing with the MiSeq platform is not susceptible to homopolymeric miscalls because it does not rely on the linearity of the relationship between the number of consecutive identical nucleotides and signal peak. Using the MiSeq sequencing platform, 123,822 raw reads were sampled from 678,702 total paired-end reads of the 317 nucleotide control amplicon that passed Illumina quality filters. From these, 2,710 Primer ID consensus sequences were generated. Using the MiSeq platform to sequence the same set of controls yielded an error rate of 1.2/10,000 bases (95% CI: 0.59-2.4) compared to 24/10,000 bases (95% CI: 18-32) using raw sequence data (Appendix 4.1). All errors were substitutions, and no difference was observed within homopolymeric regions (1.1 errors/10,000 nucleotides, 95% CI: 0.51-2.5). However, higher error rates were observed in the downstream than in the upstream paired-end sequence (rate ratio: 2.9, 95% CI: 1.7-5.0). Within the downstream sequence, errors increased over the run 2.6 times per 100 nucleotides sequenced (95% CI: 1.3-5.0). Positional errors associated with the MiSeq platform are shown in Figure 4.2B.

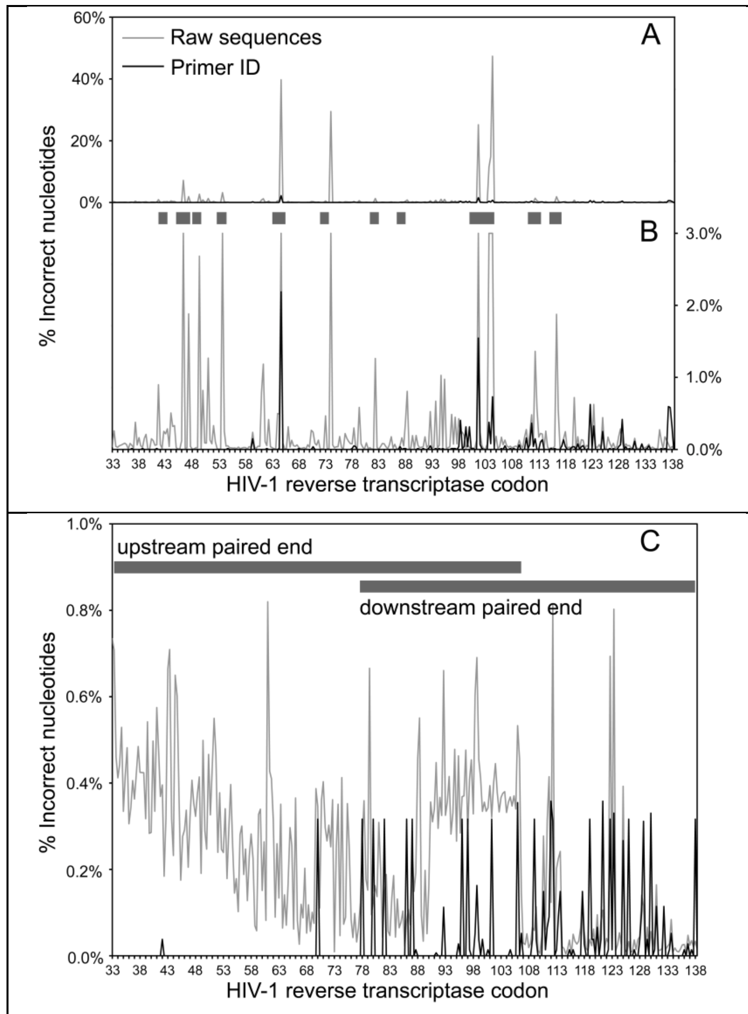

Figure 4.2. Percent incorrect nucleotide calls for each queried position of control sequence. (A) and (B) Positional errors for sequences read by the 454 Junior deep sequencing platform, with (B) enlarged to visualize very low error frequency of Primer ID corrected sequencing. Homopolymeric tracts are highlighted by dark gray bars. (C) Positional errors for sequences read using the Illumina MiSeq deep sequencing platform. The read length of each paired end is highlighted by dark gray bars. Error frequency was calculated for each platform across all dilutions submitted for deep sequencing.

4.3.2 PREVALENCE OF PRE-THERAPY DRUG RESISTANCE

Of 331 otherwise eligible participants in the UCHCC, 184 (56%) had an archived pre-therapy sample. Most were chronically infected, with median 254 (inter-quartile range (IQR): 95-398) CD4+ T cells/µL and 4.8 (IQR: 4.2-5.3) log_{10} HIV-1 RNA copies/mL prior to therapy (Table 4.1). The sequence analysis for these 184 participants was based on the identification of resistance mutations to any reverse transcriptase inhibitor (RTI), or specifically to a nucleoside/tide RT inhibitor (NRTI) or

a non-nucleoside RT inhibitor (NNRTI), with the resistance mutations defined by the 2009 list of surveillance drug resistance mutations [189]. We excluded homopolymer-associated positions from these prevalence estimates to allow for later comparison to the 454 platform. Based on sequencing of a bulk PCR product, 2.7% of participants (5/184, 95% CI: 0.89-6.2%) had an RTI resistance-associated mutation. NRTI-associated resistance was most common being present in 2.2% of participants (4/184, 95% CI: 0.60-5.5%), while 1.1% of participants (2/184, 95% CI: 0.13-3.4%) had an NNRTI resistance-associated mutation. The RTI resistance mutations detected by bulk sequencing in these five participants are shown in Figure 4.3A.

Table 4.1. Baseline characteristics of clinical participants.

Characteristic	All eligible N=331		Sample available N=184	
Gender, n (%)				
Female	77	(23%)	42	(23%)
Male	254	(77%)	142	(77%)
Race, n (%)				
Black	181	(55%)	98	(53%)
White	92	(28%)	51	(28%)
Other	58	(17%)	35	(19%)
Age, median (IQR)*	38	(31-46)	38	(31-47)
HIV risk group				
MSM, n (%)	144	(44%)	80	(43%)
IDU, n (%)	29	(8.8%)	15	(8.1%)
Heterosexual, n (%)	196	(59%)	110	(60%)
Year of 1st therapy, n (%)				
1999-2001	104	(31)	38	(21%)
2002-2004	99	(30%)	64	(35%)
2005-2007	83	(25%)	54	(29%)
>2007	45	(14%)	28	(15%)
1st regimen, n (%)				
NRTI only	45	(14%)	22	(12%)
NVP	22	(6.7%)	9	(4.9%)
EFV	264	(80%)	153	(83%)
HIV-1 RNA log_{10} copies/mL, median (IQR)	4.8	(4.3-5.4)	4.8	(4.2-5.3)
CD4+T cells/μL, median (IQR)	205	(54-357)	254	(95-398)

IQR=interquartile range; MSM=men who have sex with men; IDU=injection drug use; NRTI=nucleoside reverse transcriptase inhibitor; NVP=nevirapine; EFV=efavirenz. *Age is calculated using the date of antiretroviral therapy initiation.

Across 184 participants, >10^6 raw sequences were generated using the 454 sequencing platform, 73% (746,809) of which were >300 nucleotides long. If we relied on raw sequence data, nearly all participants had evidence of an RT inhibitor resistance mutation (148/184, 80%), but 75% of these mutations occurred among <1% of sequences within a population (Table 4.2). Excluding these very low abundance mutations in the analysis of the raw reads, 21% of participants (38/184, 95% CI: 15-27%) had any RT inhibitor resistance mutations, including 18% (34/184, 95% CI: 13-25%) and 7.1% (13/184, 95% CI: 3.8-12%) with an NRTI or NNRTI resistance mutation. However, these estimates overlook the effects of allelic bias

during PCR amplification, PCR resampling, and potential hotspots for error incorporation.

Table 4.2. Resistance mutations within clinical participant samples estimated using raw or Primer ID sequences.

	454 FLX sequences*			
	Raw sequences only		Raw & Primer ID sequences	
	No. of participants	(% w/ minority variants)‡	No. of participants	(% w/ minority variants)‡
Not homopolymer-associated†				
NRTI				
M41L	62	(85%)	7	(0%)
T69D	16	(81%)	3	(0%)
K70R	14	(100%)	1	(100%)
V75A	17	(94%)	0	(0%)
F77L	50	(90%)	6	(33%)
M184I	20	(100%)	1	(100%)
L210W	25	(92%)	3	(0%)
T215Y	12	(92%)	1	(0%)
T215F	5	(80%)	0	(0%)
T215I	40	(100%)	1	(100%)
T215S	0	(0%)	2	(50%)
T215D	7	(100%)	1	(0%)
NNRTI				
V106A	10	(90%)	0	(0%)
Y181C	18	(89%)	1	(0%)
Y188C	12	(100%)	1	(100%)
G190A	27	(96%)	5	(40%)
G190S	14	(86%)	1	(0%)
G190E	23	(91%)	2	(50%)
Homopolymer-associated†				
NRTI				
K65R	62	(98%)	0	(0%)
D67N	150	(44%)	20	(0%)
D67G	49	(100%)	1	(100%)
D67E	55	(100%)	1	(0%)
L74V	38	(87%)	4	(25%)
L74I	55	(98%)	3	(67%)
K219Q	109	(85%)	9	(11%)
K219E	18	(100%)	1	(100%)
K219N	18	(100%)	1	(100%)
K219R	24	(96%)	0	(0%)
NNRTI				
L100I	43	(95%)	0	(0%)
K101E	63	(76%)	7	(57%)
K103N	100	(91%)	2	(0%)
K103S	15	(100%)	1	(0%)

NRTI=nucleoside reverse transcriptase inhibitor; NNRTI=non-NRTI. *Primer ID used to generate consensus sequences from ≥3 raw sequences sharing a Primer ID. †Positions near homopolymeric runs, ≥4 consecutive identical nucleotides plus 2 flanking nucleotides. ‡Minority variants detected among <1% of raw sequences.

We next used the Primer IDs to form consensus sequences from the reads that represented PCR resampling. From a median 1,475 (IQR: 598-2,471) raw sequences per subject, a median 41 (IQR: 18-76) consensus sequences were constructed per subject corresponding to an average sequencing depth for reliable detection of about 7% (IQR: 4-17%). The large reduction in useable reads from the raw reads to the consensus sequences is a function of removing PCR resampling with Primer ID tagging to reveal the actual number of templates sampled. We observed that only 5-15% of the RNA templates added to the cDNA reactions resulted in consensus sequences, indicating either inefficient cDNA priming and/or extension, or inefficient inclusion of cDNA products into the PCR.

In our first analysis using Primer ID, a resistance mutation was considered if it appeared in any consensus sequence created using Primer ID, even if it appeared in a single consensus sequence (Figure 4.1C, method 1). A total of 14% (26/184, 95% CI: 9.4-20%) of participants had RTI resistance-associated mutations among Primer ID consensus sequences using the 454 platform, including 11% (20/184, 95% CI: 6.8-16%) of participants with NRTI resistance, and 4.9% (9/184, 95% CI: 2.3-9.1%) of participants with NNRTI resistance. All of the RTI resistance mutations observed by bulk sequencing were also observed in the Primer ID consensus sequences (Figure 4.3B). Conversely, using Primer ID consensus sequences rather than raw sequences resulted in a 33% reduction in the number of participants where a resistance mutation was observed (21% versus 14%), even after using a conservative (but arbitrary) 1% cut-off for mutations in the raw reads.

A

	Wild type	M	T	K	F	***Y***	M	***Y***	***G***	L	T
	RT codon	41	69	70	77	***181***	184	***188***	***190***	210	215
		Amino acid change, bulk PCR sequencing									
Subject ID	AACC	L	–	–	–	***C***	–	–	***S***	W	Y/S/D
	ATAT	–	D	–	–	–	–	–	–	–	–
	CAAT	L	–	–	–	–	–	–	–	–	–
	CATG	–	–	–	L	–	–	–	–	–	–
	CCGA	–	–	–	–	–	–	–	***A***	–	–

B

	Wild type	M	T	K	F	***Y***	M	***Y***	***G***	L	T
	RT codon	41	69	70	77	***181***	184	***188***	***190***	210	215
		Amino acid change, 454 FLX deep sequencing									
Subject ID	AACC	L	–	–	–	***C***	–	–	***S***	W	Y/S/D
	ATAT	–	D	–	–	–	–	–	–	–	–
	CAAT	L	–	–	–	–	–	–	–	–	–
	CATG	–	–	–	L	–	–	–	–	–	–
	CCGA	–	–	–	–	–	–	–	***A***	–	–
	AACA	–	–	–	–	–	–	–	–	–	(I)
	AACT	L	–	–	–	–	–	–	–	–	–
	AAGG	L	–	–	–	–	–	–	–	–	–
	AATC	(L)	–	–	(L)	–	–	–	–	–	–
	ACAC	–	–	–	–	–	(I)	–	–	–	–
	ACAG	–	–	–	–	–	–	***(C)***	–	–	–
	AGTC	–	–	–	–	–	–	–	***(A)***	–	–
	ATAA	–	(D)	–	–	–	–	–	–	–	–
	ATGC	L	–	–	–	–	–	–	–	–	–
	ATGT	–	–	–	–	–	–	–	***(A)***	–	–
	ATTC	(L)	–	–	–	–	–	–	–	–	–
	CAGG	–	–	–	–	–	–	–	***(E)***	–	–
	CATT	–	–	–	(L)	–	–	–	–	–	–
	CCAC	–	D	(R)	–	–	–	–	–	–	–
	CCAG	–	–	–	(L)	–	–	–	–	–	–
	CCAT	–	–	–	–	–	–	–	***(A)***	–	–
	CCGC	–	–	–	(L)	–	–	–	–	–	–
	CCGG	–	–	–	–	–	–	–	–	–	S
	CCTC	–	–	–	–	–	–	–	–	W	–
	CCTT	–	–	–	–	–	–	–	–	(W)	–
	CGAA	–	–	–	–	–	–	–	***(E)***	–	–

Figure 4.3. Participants with pre-existing RT inhibitor resistance mutations. (A) RT inhibitor resistance genotype for 5 participants with mutations detected using standard sequence analysis. (B) RT inhibitor resistance genotype for 26 participants with resistance detected using the 454 FLX deep sequencing platform, corrected using Primer ID. Mutations associated with a single Primer ID consensus sequence within a subject sample are shown in parentheses. RT codons and mutations associated with NNRTI resistance are highlighted in bold italic type, while RT codons and mutations associated with NRTI resistance are shown in standard type. Only RT codons outside of homopolymeric influence were included in this analysis. Lack of resistance for a particular RT codon is indicated by a dash.

The frequency of single mutations of any type in the data set of Primer ID consensus sequences was four as high as expected given the error rate determined using the plasmid sequences. Thus, in most cases the call of a resistance mutation based on a single observation was likely accurate. However, in a second, more

conservative analysis, only those resistance mutations that appeared in at least two consensus sequences were counted (Figure 4.1C, method 2). When only multiple (i.e. 2 or more) within-subject observations of a specific resistance mutation were included, the prevalence of RTI resistance among these 184 participants was 6.0% (11/184, 95% CI: 3.0-10%), representing 6 additional participants over the 5 who were also identified using bulk sequence analysis (Figure 4.3B).

The preceding analysis did not include the homopolymeric regions, and we carried out a separate analysis to see what influence they would have on calls of drug resistance mutations. We found that only 4 (2%) out of 184 participants sequenced using 454 had a predominant “AAG” (a wild type lys codon) at RT codon 65, and 11 (6%) participants had an “AAG” at RT codon 65 as a minority variant with abundance ranging from 1% to 24%. No evidence of K65R was found. If other homopolymeric positions were included, an additional 24 participants would have been classified as RTI resistant using the 454 data with consensus sequences, raising the overall prevalence to 27% (50/184). Some resistance calls at homopolymeric positions were also seen by bulk sequence analysis (in three participants) and were unlikely due to homopolymeric error given their high abundance. In contrast, homopolymer-associated resistance mutations detected solely by deep sequencing ranged in frequency from 0.35%-12.5% and most appeared once within a sample. Assuming these single occurrences were miscalls due to homopolymeric error, only six additional participants would be classified as having pre-existing RTI resistance (based on the mutation being on more than one consensus sequence). Thus, if homopolymer-associated positions were included, prevalence estimates would increase to 9% (17/184).

4.3.3 Relative Abundance of Resistance Mutations within Viral Populations

The use of Primer ID allows an assessment of the number of viral genomes that were actually sampled from a subject, thus allowing an assessment of both sequencing depth and the relative abundance of detected mutations from a specific clinical sample. There were 6 participants who had resistance mutations that were detected in multiple Primer ID consensus sequences but not detected by bulk sequencing. The median abundance of these mutations within the viral population in each person was 1.8% (IQR: 1.2-2.8%). There were an additional 15 participants who had a resistance mutation present in one Primer ID consensus sequence and the median abundance of these mutations was 1.7% (IQR: 0.67-1.1%). However, the estimate of abundance is significantly limited given the low number of observations of each mutation, and the ability to detect variants at even less abundance is limited by low template utilization, as revealed using the Primer ID. This phenomenon is highlighted in Figure 4.4, where the majority of low abundance resistance was detected on a single Primer ID consensus sequence within a subject sample.

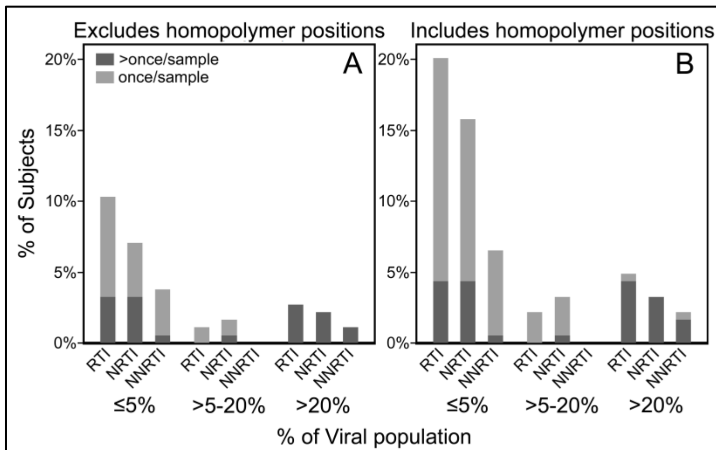

Figure 4.4. Prevalence and frequency of pre-existing RT inhibitor resistance among participants. The left-hand panel excludes homopolymeric tracts (≥4 consecutive, identical nucleotides plus 2 flanking nucleotides). Resistance was defined using the surveillance drug resistance mutations [189].

Using Primer ID, few participants with RTI resistance had evidence of multiple resistance mutations. Of 21 participants with minority drug resistance, only 2 (10%) had more than one drug resistance mutation, each occurring on separate Primer ID consensus sequences at very low frequency. Of 5 participants with a majority drug resistant population, only 1 (20%) had multiple resistance mutations. This subject had extensive drug resistance that was also revealed by bulk sequence analysis: Y181C, G190S, and L210W appeared with T215Y, T215S, or T215D among 79% (26/29), 10% (3/29), or 10% (3/29) of consensus sequences, respectively, while M41L was linked to homopolymer-associated L74V and K101E in 94% (29/31) of consensus sequences. Together, this suggests that this subject was initially infected with a variant carrying M41L, L74V, K101E, Y181C, G190S, L210W, and T215Y mutations, with the virus slowly reverting at codons 74 and 215.

4.3.4 COMPARISON OF DEEP SEQUENCING PLATFORMS IN A CLINICAL SETTING

Sequences spanning HIV-1 RT codons 34-74 and 111-139 (HXB2 nucleotides 2648-2770 and 2878-2964) were determined for 19 of 184 participants using the Illumina MiSeq platform. Based on previous analyses of the data from the 454 FLX sequencing platform, we selected participants who had the most consensus sequences constructed from ≥3 raw sequences sharing the same Primer ID (median 203 [IQR: 168-247] consensus sequences), indicating that these samples had the highest level of genomes incorporated into the cDNA/PCR step. Using the MiSeq platform, a median 273 (IQR: 192-583) consensus sequences were constructed from ≥5 raw sequences sharing a Primer ID (median 29,743 (IQR): 24,686-33,086 raw sequences). For 17 of the 19 participants, the number of consensus sequences generated using the MiSeq platform was comparable to those obtained using the 454 FLX platform despite a nearly 10-fold increase in raw sequences. For the remaining 2 participants, the number of consensus sequences increased 6- or 12-fold. While

every effort was made to sequence the same amplicons submitted for 454 sequencing, RNA extraction was repeated for these 2 participants since their cDNA and amplicons were previously exhausted. In these 2 cases, high viral titers (6.6-7.9 log_{10} HIV-1 RNA copies/mL) and dilution error could explain the discrepancy in apparent higher template utilization.

A total of 108 amino acid changes were observed in the RT coding region using both sequencing platforms, with 74 mutations detected using the 454 platform and 82 using the MiSeq platform. About half (48/108) were detected by both sequencing platforms, 25% (26/108) were detected by the 454 platform alone, and 33% (34/108) were detected solely by the Illumina MiSeq platform. Nearly 72% (43/60) of mutations detected by a single sequencing platform occurred on a single Primer ID consensus sequence, suggesting these mutations were either the result of method error or of stochastic sampling of rare variants. All 34 variants detected solely by the Illumina MiSeq platform were within the downstream paired end sequence, which was revealed as an error hotspot by the control experiments (Figure 4.2C). Conversely, 38% (10/26) of variants detected by the 454 platform alone were associated with homopolymeric regions, but most of the variants outside of homopolymeric regions occurred once (10/16, 62%) and could be due to a low number of templates.

4.4 Discussion

Deep sequencing methods are subject to bias introduced by PCR amplification, and those methods that allow consecutive nucleotide additions in a homopolymeric run are also vulnerable to erroneous calls in or near these runs [369]. Here, an alternative deep sequencing method that tags a single viral template with a unique Primer ID prior to PCR [26] was used to estimate the prevalence of pre-

existing RTI resistance within a clinical population initiating care for HIV-1. Among 184 participants, up to 14% had evidence of RTI resistance, compared to 2.7% detection by sequencing of a bulk PCR product. An even more conservative use of the deep sequencing data based on making calls only if the mutation was associated with at least two consensus sequences gave an RTI resistance mutation detection rate of 6.0%, still more than a 2-fold increase over that seen by sequencing of a bulk PCR product, and these estimates did not include an analysis of homopolymeric regions that are susceptible to especially high error rates using the 454 sequencing platform.

Prevalence estimates must be critically interpreted since the value can be inflated due to several intrinsic errors in the sequencing methodology, not all of which can be corrected by Primer ID. The 454 platform control experiments demonstrated nearly 4-fold higher error rates within homopolymeric regions compared to heteropolymeric regions despite the use of 3 or more raw sequences with the same Primer ID to create a consensus sequence. Intractable homopolymeric errors argue against using sequencing platforms that are subject to these errors to estimate the prevalence of minority variants, especially those associated with homopolymeric regions such as variants with K103N and K65R [346, 366, 370]. The Illumina MiSeq platform, which does not rely on the incorporation of multiple nucleotides at a homopolymeric stretch, eliminated homopolymer-associated errors in control experiments. However, this system has its own set of limitations including the accumulation of errors over the sequencing run [326], which was consistent with our own control experiments, and poor discrimination of highly similar sequences [371]. When Illumina was used to sequence samples from a subset of participants and compared to the 454 sequencing platform, concordance between the two platforms

was only 50% across all queried positions, suggesting either substantial sequencing error or stochastic sampling, particularly associated with very low abundance variants. There is some evidence of homopolymer-associated sequencing error, since nearly 40% of mutations detected only by the 454 platform were near or within these sites. It is also likely that many of the mutations detected by the MiSeq platform alone were the result of sequencing error, since all of these mutations were within the downstream paired-end which is associated with a 3-fold increase in error. However, it is not possible to rule out stochastic sampling of the viral population as the source of the discrepancy in most cases given the limited template usage revealed by the use of Primer ID.

Our analysis of clinical subject samples was clearly limited by the number of templates we sampled, and if sufficient numbers of templates had been available (i.e. enough to give 1,000 or more consensus sequences) we could have queried down to the 0.1% to 0.5% range, below which residual method error still confounds the analysis. While limited template utilization was a problem in our analysis of these samples, it was the use of Primer ID that revealed the extent of template utilization and allowed us to estimate the quality of sampling. Alternatively, if we had relied on the raw reads with an arbitrary cut-off (1%), we would have not only overestimated the prevalence of RT inhibitor resistance, but we would have also erroneously concluded that our sampling depth was much higher for these samples given the number of raw reads that passed quality filters (median >2,000), and our estimates of the frequency of resistance mutations in the viral population would have been skewed upwards by nearly 20% compared to Primer ID.

Even after correction with Primer ID, including all resistance mutations in estimates, i.e. even those that appear in only one Primer ID consensus sequence,

may fail to correct for errors that are introduced during cDNA synthesis, that occur in the earliest cycles of PCR amplification, or that are homopolymer-associated. Unfortunately, downstream data filtering with Primer ID cannot account for the first two of these biases, but control experiments did demonstrate a substantial reduction in errors within homopolymeric regions. Sequencing errors within the Primer ID itself cannot be ruled out, either, and these errors may be even more likely if the Primer ID itself contains a homopolymeric sequence. In the worst-case scenario, a viral genome is linked to a homopolymeric Primer ID and subsequently oversampled, such as might occur when the number of input templates is low, and thus the number of reads of each Primer ID is high. Since the original Primer ID itself contains a homopolymeric sequence, it is more likely to be misread repeatedly and in the same way by the 454 sequencing platform. In this manner, more than one Primer ID may be linked to the same viral genome, and these would be counted as separate viral genomes when collapsed into separate consensus sequences. Most such Primer IDs are unlikely to be abundant enough to be included in consensus sequence assembly. We assessed this type of error by building a tree of the Primer ID sequences themselves. We found no evidence of this type of oversampling in this dataset, although this type of monitoring is likely to be an important feature of using Primer ID. Finally, Primer ID may also fail when there are ties in nucleotide calls at a given position (ambiguity) among resampled raw sequences, thus making it impossible to infer the “real” viral sequence. This scenario is more likely to occur when consensus sequences are constructed from a low number of resampled raw sequences. Among 184 subject samples, each Primer ID consensus sequence was constructed from a median 12 (IQR: 6-24) resampled raw sequences, and therefore, this scenario cannot be excluded.

Despite its limitations, Primer ID offers an opportunity to make inferences about changes within the viral population when multiple resistance mutations are present, assuming that each unique Primer ID represents an individual viral genome. Only three participants had evidence of multiple resistance mutations on non-homopolymer-associated codons, one of whom was identified by bulk sequence analysis. Two participants had multiple, low frequency NRTI resistance mutations on separate Primer ID consensus sequences, which could indicate past NRTI exposure followed by the reappearance of wild type from the participants' reservoirs. The remaining subject had multiple, linked resistance mutations (M41L+L74V+K101E and Y181C+G190S+T215Y/D/S) that predominated the Primer ID consensus sequences with evidence of reversion at codons 74 and 215. Previous studies have shown that K101E+G190S reduce fitness compared to wild type virus in the absence of antiretroviral therapy, but that the addition of M41L+T215Y or L74V in particular improves fitness without reducing NNRTI resistance [372, 373]. In this subject it appears that T215Y is reverting more rapidly than L74V given the higher frequency of sequences with T215 revertant mutations compared to L74 (21% vs. 6%), but we are limited in our conclusions since these regions of RT were independently amplified and sequenced and thus for these different amplicons cannot be linked.

Many studies, including a recent systematic review [18], have linked minority pre-therapy NNRTI resistance with an increased risk of virologic failure. Despite this evidence, questions still remain surrounding the clinical importance of minority drug resistant variants, particularly with respect to defining a specific abundance threshold at which resistance begins to affect response to combination therapy [364]. Before any particular cut-off for clinical significance can be determined, the drug resistant viral population must first be measured as accurately as possible. The most

promising method with potential to move beyond the research setting, ultra deep sequencing, still suffers from multiple sources of error that are inherent in this method. In this study, these errors and PCR resampling were addressed using the Primer ID, which showed a 30-fold reduction in error rates over raw sequence analyses and which, despite limited viral template usage in clinical samples, still revealed additional participants with pre-therapy resistance. As important, the use of Primer ID reveals the number of templates that were actually sampled thus providing an accurate assessment of the quality of the sampling depth, an essential piece of information when evaluating the meaning of the detection of rare variants.

CHAPTER 5 VIROLOGIC CONSEQUENSES OF PRE-EXISTING LOW ABUNDANCE RESISTANCE AMONG PATIENTS INITIATING WITH A REVERSE TRANSCRIPTASE INHIBITOR REGIMEN

5.1 INTRODUCTION

Accurate pre-therapy resistance profiles are critical to initial treatment selection [363]; however, standard sequencing fails to reliably sample viral variants <20% of the population [336]. This limitation has implications for chronically-infected patients with resistant virus that has faded to levels below detection, yet still remains clinically relevant [128]. Ultra deep sequencing (UDS) is capable of achieving sampling depths <1% of viral populations [23], but is limited by a PCR step that biases estimates upward through enzyme-mediated errors, differential amplification, and sequence resampling [20]. Some UDS platforms are additionally vulnerable to miscalls near homopolymeric nucleotide regions [24]. These artifacts must be distinguished from low level HIV resistance that may affect treatment [18].

In this study, to minimize errors, each viral template was tagged by a random 8-nucleotide sequence during cDNA synthesis, allowing identification of specific transcripts that were subsequently amplified. Amplified sequences sharing the same 8-nucleotide Primer ID originate from the same genome so that concordant nucleotide changes from a consensus sequence are more likely to be actual mutations and discordant nucleotide changes are more likely to be errors. Creating consensus sequences from ≥3 sequences sharing the same Primer ID allows for fewer miscalls and estimation of input templates [26]. In patients receiving highly active antiretroviral therapy (HAART), this method was used to estimate the

proportion who received less than 3 fully-active antiretrovirals (ARV) and to examine the association between suboptimal HAART and virologic failure (VF).

5.2 Methods

Patients enrolled in the University of North Carolina Center for AIDS Research HIV Clinical Cohort Study (UCHCC) were included if they: (1) provided written, informed consent for enrollment in the cohort study and use of their specimens for research purposes; (2) initiated HAART after December 31, 1999; (3) with either ≥2 nucleoside(tide) reverse transcriptase inhibitors (NRTI) plus an non-NRTI or ≥3 NRTI; and (4) had at least one recorded pre-therapy HIV-1 RNA level; and (5) had an archived sample available for sequencing. UCHCC is an ongoing clinical cohort study enrolling adults seeking HIV care at UNC. UCHCC maintains an electronic database of patient information and repository of patient plasma samples [339]. This study was approved by the UNC Institutional Review Board.

HIV-1 RNA was extracted from 140 µL of plasma, and concentrated if HIV-1 RNA was <4.5 $\log_{10}$ copies/mL. One-third of the RNA was added to two cDNA reactions [26] using HIV-1 reverse transcriptase (RT) specific primers: (1) 5'-GCCTTGCCAGCACGCTCACAGCTGGCA-NNNN-CGNNNNNNNNTC-ACATTGTACTGATATCTAATYCCTGGTG-3' complimentary to HXB2 nucleotides 2965-2992 (fragment 1) [346]; and (2) 5'-GCCTTGCCAGCACGCTCACAGCTGGCA-NNNN-CGNNNNNNNNTC-CACTATAGGCTGTACTGTCCATTTATC-3' complimentary to HXB2 nucleotides 3258-3284 (fragment 2) [346]. Primers featured a patient barcode (NNNN) [22], a random template specific Primer ID (CGNNNNNNNNTC) [26], and a PCR primer with a *PvuII* cut site (Figure 5.1A).

DNA was amplified using primers: (1) 5'-GGCCATTGACAGAAGAAAAAATAAAAGC-3', sense, fragment 1 [346]; (2) 5'-

CACCAGGRATTAGATATCAGTACAATGT-3’, sense, fragment 2 [346]; (3) 5’-GCCTTGCCAGCACGCTCACAG-3’, antisense, round 1; and (4) 5’-CCAGCACGCTCACAGCTGGCA-3’, antisense, round 2. PCR conditions included 30 cycles with 67°C (fragment 1) or 63°C (fragment 2) annealing temperatures (Phusion High Fidelity Hot Start II, Thermo Scientific, Waltham, MA). Amplicons were pooled in equimolar concentration and *PvuII* digested. Pools were gel purified, quantified, and submitted for 454 GS FLX sequencing (Roche, Indianapolis, IN).

UDS sequences <300 nucleotides and those with ambiguous patient specific barcodes or Primer ID were discarded. Remaining sequences were partitioned by patient barcode. Sequences with template specific Primer ID that occurred <3 times were discarded as discrepancies between sequences could not be adjudicated, and majority-rules consensus sequences were constructed from remaining sequences sharing the same Primer ID. These were input into the Stanford HIV Drug Resistance Database (HIVdb) for interpretation [171]. To increase confidence of calls near homopolymeric regions, mutations on HIV-1 RT codons 40, 41, 44, 62, 65, 66, 67, 69, 70, 71, 74, 77, 98, 100, 101, 103, 115, 116, 118, 151, 188, 215, 219, 221, and 227 were considered genuine if they were associated with more than one Primer ID within a patient sample (Figure 1A). In sensitivity analyses, these criteria were extended to HIV-1 RT codons outside of homopolymeric regions. The number of active antiretroviral (ARV) in the initial regimen was calculated, with HIVdb scores ≥15 considered resistant.

Bulk sequences were obtained using HIV GenoSure (LabCorp, Research Triangle Park, NC) or by in-house bulk sequencing using 2nd round PCR primers. The majority of bulk sequencing was performed retrospectively. All patients with

prospective results available (N=37/184, 20%) were prescribed fully-active HAART as predicted by HIVdb interpretation of bulk sequences [171].

Patients were followed from therapy initiation until the HIV-1 RNA level before the earliest of the following: (1) switch to non-reverse transcriptase inhibitor (non-RTI) HAART; (2) discontinuation of HAART for >2 weeks; or (3) August 17, 2012. Patients experienced VF if they did not achieve 0.5 $\log_{10}$ HIV-1 RNA copies/mL decrease from baseline in the first 6 months of therapy, or after 6 months, if they had an HIV-1 RNA level >400 copies/mL.

The combined association of NNRTI or NRTI resistance to all RTI regimens received with time-to-VF was examined. Drug resistance mutations were categorized as follows: (1) overall (<3 active ARV vs. ≥3 active ARV); (2) by relative abundance (minority or dominant vs. ≥3 active ARV); and (3) by absolute copy number of RTI resistant sequences, calculated using the proportion of resistant consensus sequences and sample viral load. Finally, the receipt of <3 active ARV was categorized by sequencing detection method (by UDS only, or by both bulk sequencing and UDS vs. ≥3 active ARV).

Causal diagram analysis was used to identify adjustment variables, which consisted of gender, race (white vs. non-white), men who have sex with men (MSM), and age at therapy initiation (modeled using restricted cubic splines with knots at the quartiles of the population distribution). To account for lag between resistance measurement and the beginning of therapy, a continuous variable for months from sample draw to therapy initiation was created and included as an interaction term with resistance.

Cumulative incidence curves for time-to-VF were stratified by key variables and compared using log-rank tests. Proportional hazards models, unadjusted and

adjusted and stratified by months from UDS sample draw to therapy initiation, were fit for each resistance variable. Hazard ratios (HR) and 95% confidence intervals (CI) compared the hazard (instantaneous rate) of VF among patients with resistance to their initial regimen, measured at the time of therapy initiation, to those initiating with all active ARV. The proportional hazards assumption was evaluated by inspection of log[-log(survival)] curves.

5.3 RESULTS

As of January 2010, 768 (37%, N=768/2076) UCHCC patients initiated therapy after 2000, 367 started with an RTI regimen, and 331 had ≥1 reported pre-therapy HIV-1 RNA level. Of these, 184 (56%) patients had available pre-therapy samples. Median age at first therapy was 38 years (inter-quartile range (IQR): 31-47). Patients were mostly male (77%), one-half African American (53%), and 43% and 8% of patients were MSM or had a history of injection drug use (IDU). Patients had median baseline 254 CD4+ T cells/μL (IQR: 95-398) and 4.8 $\log_{10}$ HIV-1 RNA copies/mL (IQR: 4.2-5.3).

Most patients initially received efavirenz (EFV) (N=153, 83%), 9 received nevirapine (5%), and 22 (12%) received exclusively NRTI. Of NNRTI initiating patients, 78 (48%) were coadministered emtricitabine and tenofovir, and 58 (36%) received lamivudine (3TC) and zidovudine (AZT). Of NRTI initiating patients, 20 (91%) were given abacavir, 3TC, and AZT. Patients received RTI regimens for a median 13 months (IQR: 3.3-39). Nearly 38% (N=70) of patients had not experienced VF and remained on RTI regimens at the close of follow-up; 23% (N=43) and 14% (N=26) of patients were censored because they discontinued therapy for >2 weeks or switched to a non-RTI.

UDS and bulk sequencing data were obtained for 182 and 180 patients; combined, 181 had both. Overall, 10% (N=19/181, 95% CI: 6-15%) of patients received a regimen with fewer than three active RTI, including 3 patients identified by both technologies. These 3 patients received regimens containing 3TC, AZT, and EFV and had evidence of the following mutations: M41L, L74V, K101E, Y181C, G190S, L210W, and T215Y/D/S; K103N/S and G190A; and M41L alone. An additional 9% (N=16/181; 95%CI: 5-14%) of patients had resistance detected by UDS alone, about half of which was NNRTI-associated (N=9/16). The most common mutations detected were G190A/S/E/V (N=6) M41L (N=3), and D67N (N=3). The characteristics of patients who received ≥3 active vs. <3 active ARV are compared in Table 5.1.

Table 5.1. Baseline characteristics of patient population by number of active drugs.

	<3 active RTI* N=19		≥3 active RTI* N=162	
Women, n (%)	3	(16%)	38	(23%)
Age at 1st regimen, median (IQR)	41	(32-47)	38	(31-47)
African American, n (%)	10	(53%)	86	(53%)
MSM, n (%)	9	(47%)	70	(43%)
IDU, n (%)	1	(5.3%)	14	(8.6%)
1st regimen, n (%)				
NRTI only	3	(16%)	19	(12%)
EFV based	14	(74%)	137	(85%)
NVP based	2	(10%)	6	(3.7%)
Retrospective genotype, n (%)	17	(89%)	127	(78%)
Deep sequencing to 1st regimen, n (%)				
Same day	13	(68%)	99	(61%)
1-90 days	6	(32%)	42	(26%)
90+ days	9	(0.0%)	21	(13%)
$\log_{10}$ HIV-1 RNA co/mL, median (IQR)	5.0	4.6-5.6	4.8	(4.1-5.2)
CD4+ T cells/μL, median (IQR)	150	(33-306)	257	(108-402)
Months follow-up, median (IQR)	18	(7.3-50)	12	(1.9-38)

RTI=reverse transcriptase inhibitor; IQR=inter-quartile range; MSM=men who have sex with men; IDU=injection drug use; NRTI=nucleoside reverse transcriptase inhibitor; EFV=efavirenz; NVP=nevirapine *Mutations on HIV-1 reverse transcriptase codons 40, 41, 44, 62, 65, 66, 67, 69, 70, 71, 74, 77, 98, 100, 101, 103, 115, 116, 118, 151, 188, 215, 219, 221, or 227 were considered genuine only if they occurred >1 time within a patient sample. The number of active RTI per regimen was predicted using the Stanford HIV drug resistance database to interpret mutations detected by bulk PCR sequence analysis and/or ultra deep sequencing [171].

Nearly one-fourth of patients (N=45) experienced VF within a median 60 months (IQR: 17->142). Patients receiving sub-optimal regimens trended toward shorter time-to-VF than patients receiving 3 or more active RTI (HR: 1.8, 95% CI: 0.87-3.7, p=0.1) (Figure 5.1B). Shorter failure times were driven largely by bulk sequencing predicted resistance; 2 of 3 patients, both of whom received inactive EFV, failed within 7.6 months (Figure 5.1C). Patients taking fewer than 3 active RTI predicted by UDS alone experienced VF at rates slightly higher than but not statistically different from those without resistance (HR: 1.6, 95% CI: 0.71-3.6, p=0.3), but the power to detect a difference was limited by the small number of patients with resistance.

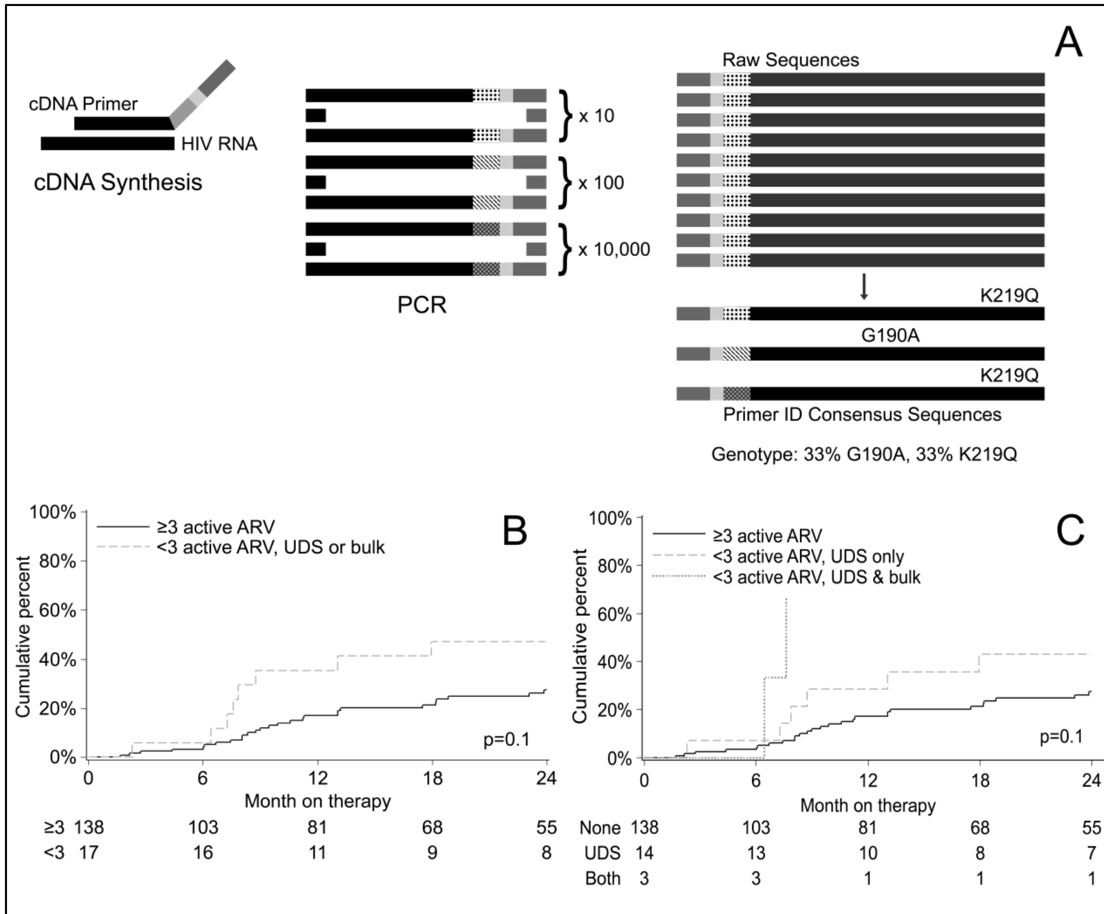

Figure 5.1. Activity of early therapy predicted using Primer ID corrected deep sequencing genotype and time-to-virologic failure. Panel A illustrates the Primer ID correction method [26]. Individual HIV-1 RNA within a patient sample are tagged during cDNA synthesis with a randomized Primer ID sequence, a patient barcode and a PCR primer sequence. Sequences linked to unique Primer ID (dots, diagonal lines, hash) may be differentially amplified during PCR. Consensus sequences are constructed from >3 raw sequences sharing the same Primer ID within a sample and interpreted using the Stanford drug resistance database [171]. The association between receiving fewer than three active reverse transcriptase inhibitors over follow-up and time-to-virologic failure is shown in panels B and C. ARV=antiretroviral; UDS=ultra deep sequencing; bulk=bulk PCR sequencing.

The adjusted analysis produced similar results (Table 5.2). Shorter time to VF was associated with more abundant resistant variants, with mutations present among ≥20% of the total viral population having a larger effect on time-to-VF (HR: 2.2, 95% CI: 0.77-6.5, p=0.1). Failure times were shorter for patients with as few as 2,000 copies/mL of resistant virus. Using stricter interpretations of UDS data drove estimates further toward the null (Appendix 5.1).

Table 5.2. Adjusted hazard ratios for time-to-virologic failure and receiving suboptimal therapy.

				Adjusted estimates		
Resistance to early therapy*	n	VF	p-m	HR†	95% CI	95%CLR
Predicted resistance						
<3 inactive RTI, Bulk or UDS	19	9	97	1.9	0.90-4.1	4.5
≥3 active RTI	162	35	619	1.0		
Predicted resistance						
<3 inactive RTI, Bulk & UDS	3	2	14	2.9	0.65-13	19
<3 inactive RTI, UDS Only	16	7	83	1.7	0.75-4.0	5.4
≥3 active RTI	162	35	619	1.0		
% Viral population, UDS						
≥20% resistant	8	4	29	2.2	0.77-6.5	8.4
<20% resistant	11	5	68	1.7	0.64-4.5	6.9
None	163	36	635	1.0		
Copies resistant virus, UDS						
≥2,000 co/mL	14	6	56	1.6	0.65-3.8	5.9
<2,000 co/mL	168	39	676	1.0		
Copies resistant virus, UDS						
≥10,000 co/mL	9	5	38	3.5	1.0-7.1	6.9
<10,000 co/mL	173	40	694	1.0		

VF=virologic failure; p-m=person-months; CLR=confidence limit ratio; RTI=reverse transcriptase inhibitor; UDS=ultra deep sequencing; MSM, men who have sex with men. *Mutations near homopolymeric sequence regions were considered genuine if they occurred >1 time within a sample. RTI activity was predicted using the Stanford HIV drug resistance database [171]. †HR and 95% CI were adjusted for the following: (1) time-varying MSM and race; (2) gender; and (3) age at therapy initiation modeled as restricted cubic splines with knots at 25th, 50th, and 75th percentiles.

5.4 Discussion

Nearly 10% of patients received partially-active HAART, and 84% of these were not captured by standard sequencing. Although bulk sequencing was performed retrospectively for most patients, the number receiving suboptimal therapy was lower than in other cohorts with prospective testing [362]. By bulk sequencing analysis, <2% of study patients received suboptimal HAART, while 6% of patients from another cohort with clinically-available tests received suboptimal RTI combinations [362]. Additional resistance observed with UDS is consistent other groups that used more sensitive assays and reported an increase in NRTI resistance

from 3 to 13% of patients [374] or NNRTI resistance from 6% to 13% of patients [346].

In our analysis, the intra-patient distribution of pre-existing resistance was highly skewed towards the upper (≥20%) and lower (<5%) ends. Thus, a specific cut-off at which minority variants affect VF could not be determined. Others suggest the effect of RTI resistance on time-to-VF may be more a function of copies of resistant virus rather than percent abundance alone [18, 364]. In this study, the magnitude of the estimated effect of copy number exceeded the estimated effect of variants defined by a percent cut-off at the highest copy numbers. This result is not surprising, since patients with a relatively high proportion of resistant virus often had lower viral loads and may have achieved sustained viral suppression more easily.

To produce clinically relevant estimates, baseline resistance was defined in the context of the antiretroviral activity of each HAART regimen received over the course of follow-up [171], which is predictive of VF [363]. Consistent with previous bulk sequencing studies [362, 363], pre-therapy resistance was associated with a near 100% increase of the hazard of VF. Estimates were heavily influenced by NNRTI resistance detected by bulk sequencing, rather than additional resistance identified by UDS. Patients identified solely by UDS, nearly half of whom were resistant to their prescribed NRTI, experienced slightly elevated rates of VF compared to those receiving a fully active regimen, in agreement with other estimates for NRTI minority variants and VF [18]. It is possible that patients harboring minority resistance have accelerated failure times, but this study was unable to discriminate between groups with NRTI and NNRTI resistance, especially given the rigorous definition of drug resistance.

This is the first study to scale up the Primer ID based method to a large population of patients who went on to receive HAART. Primer ID allows a more accurate estimation of the viral population with UDS data [26] and may allow more refined estimation of the effects of minority drug resistant variants on virologic response given greater sampling depth and moderation of sequencing error. Predicting virologic response will likely improve by precise enumeration of minority variants actually present in the viral population pre-treatment. In order for these highly sensitive assays to complement clinical decision-making, it is important to distinguish resistance from error and to consider how resistance interacts with other factors such as viral load.

CHAPTER 6 CONCLUSION

6.1 SUMMARY OF RESULTS

A clear picture of antiretroviral resistance within the HIV-infected patient is clouded by factors that both are native to the virus, with its enormous capacity for evolution, and technology that is currently available for its measurement, which is tasked with distinguishing sequencing errors that, while rare, may occur more frequently than some mutations. Ultra deep sequencing (UDS) promises far better sampling of the viral population than standard sequencing, but UDS protocols are highly complex with multiple opportunities for error introduction. In particular, UDS still relies on a PCR amplification step to enrich rare HIV-1 sequences for an adequate sequencing signal; this step notoriously misrepresents the viral population through differential amplification of HIV-1 genomes [19]. Thus, when millions of HIV-1 sequences are returned for analysis, they are not necessarily providing a clear picture of the HIV-1 population within the individual, but rather a picture of what HIV-1 sequences were most efficiently amplified during PCR. Without accurately sampling and measuring the HIV-1 population within the individual, critical questions about the threshold at which minority drug resistant HIV-1 begins to affect treatment outcomes cannot be addressed.

One method of addressing UDS error is through the use of Primer ID. This stretch of randomized nucleotides is incorporated into the HIV-1 genome before PCR amplification so that each viral genome is tagged with a unique Primer ID sequence. Primer ID are amplified along with the HIV-1 sequence, and multiple occurrences of

the same Primer ID sequence within a sample are used to create a single consensus sequence that not only better reflects the true nucleotide sequence of that particular viral genome, but also its distribution within the sampled viral population. Others have used UDS to measure minority drug resistant HIV-1, and they have attempted to determine a threshold at which these variants begin to affect treatment outcome [18]. However, this study shows that Primer ID may be successfully applied to this problem, increasing the potential to tease out clinically meaningful thresholds that are not confounded by UDS error.

In this study, we first observed in control experiments that Primer ID virtually eliminated stochastic sequencing errors while revealing systematic errors associated with certain sequence regions. In known sequences spanning HIV-1 RT codons 34-139, the overall error rate using the 454 platform was reduced from 70/10,000 nucleotides (95% CI: 70-72) to 2.6/10,000 nucleotides (95% CI: 2.2-3.2) when raw sequences were corrected using the Primer ID. Homopolymeric sequence tracts are known to be particularly difficult to read for 454 sequencing platforms [24], and Primer ID dramatically reduced, but did not completely eliminate errors associated with these regions. Using raw sequences, we observed 216 errors/10,000 nucleotides (95% CI: 213-220) near homopolymeric sequences that were 4 or more nucleotides in length, but Primer ID reduced this rate to 6.0/10,000 nucleotides (95% CI: 4.8-7.4) within these problematic areas. In contrast, the Illumina sequencing platform is not susceptible to homopolymeric errors, but it is associated with errors toward the end of a sequencing run (ie., towards the end of the sequence read) and within the downstream paired-end sequence [326]. Using the Illumina MiSeq platform, the overall error rate for both pair-end sequences was reduced from 24/10,000 nucleotides (95% CI: 18-32) using raw sequences to 1.2/10,000

nucleotides (95% CI: 0.59-2.4) using Primer ID correction. However, while in raw sequences errors appeared more randomly distributed between the paired-ends, Primer ID revealed that residual errors were concentrated within the downstream paired-end.

We next used Primer ID corrected 454 deep sequencing of HIV-1 RT codons 34-245 to measure the amount of pre-existing RT inhibitor resistance in a population of 184 therapy naïve patients seeking HIV care in North Carolina. Although we relied on the most current list of surveillance drug resistance mutations to estimate the prevalence of transmitted drug resistance [189], we excluded those mutations near homopolymer-associated positions from our overall estimates in light of residual errors revealed by control experiments even after Primer ID correction. We also produced another set of prevalence estimates under the stricter assumption that a drug resistance mutation was genuine only if it occurred more than once within a patient sample (ie. on at least two Primer ID consensus sequences). Under the less strict interpretation, we found that nearly 14% (95% CI: 9.4-20%) of patients had one or more RT inhibitor mutations, including 11% (95% CI: 6.8-16%) with NRTI resistance and 4.9% (95% CI: 2.3-9.1%) with NNRTI resistance. Under the more conservative definition, 6.0% (95% CI: 3.0-10%) of patients had any RT inhibitor resistance; the prevalence of NRTI and NNRTI resistance was 6.0% (95% CI: 3.0-10%) and 1.6% (95% CI: 0.34-4.7%), respectively. The latter was probably overly conservative since, given the error rate estimated from controls, we observed nearly a 3-fold excess of singly occurring mutations than what we would expect by chance alone. However, even the strictest interpretation of Primer ID consensus sequences revealed an increase in resistance prevalence over standard sequencing, since only 2.7% (95% CI: 0.89-6.2%) of patients had evidence of RT inhibitor resistance by bulk

sequence analysis, including 2.2% (95% CI: 0.60-5.5%) and 1.1% of patients (95% CI: 0.13-3.4%) with NRTI or NNRTI resistance, respectively.

Resistance mutations that were detected by both standard bulk sequencing and UDS dominated the viral population, and all were present in far greater than 20% of Primer ID consensus sequences. In contrast, resistance mutations detected solely by UDS were in very low abundance (<1% to <5%), and these most often occurred on a single Primer ID within a patient sample. Our goal was to input 500 HIV-1 RNA templates into the sequencing reaction, but the median number of Primer ID consensus sequences per patient was 41 (IQR: 18-76) created from a median 12 raw sequences (IQR: 6-24) each, suggesting very limited sampling depth and low template utilization. We would not have observed this limitation if we had relied on raw sequences alone, since patients had a median 1,475 raw sequences (IQR: 598-2,471), but since we restricted our analysis to mutations outside of homopolymeric influence, it is likely that even single occurrences of resistance mutations were authentic.

When we compared all amino acid changes in a subset of 19 patients between results from Illumina MiSeq and 454, we found similar numbers of Primer ID consensus sequences for 11 of these patients despite a 9-fold increase in the number of raw sequences, while 8 had up to 9-fold more Primer ID consensus sequences using the Illumina platform. We could not explain this difference between platforms for all patients even though Illumina MiSeq outputs vastly more raw sequences compared to 454 platforms; however, 4 of 8 of these patients did have viral loads exceeding the limit of detection (>750,000 HIV-1 RNA co/mL), so the number 454 Primer ID consensus sequences could have been limited by platform throughput rather than template usage. When we compared all amino acid changes

between platform, we found that only 50% (N=48/108) were detected by both UDS platforms. Most of these (72% of 48) were detected on a single Primer ID consensus sequence. Further, up to 73% (N=19/26) of mutations detected by the 454 platform alone were near homopolymeric sequences, while all 33 mutations detected by the Illumina platform alone were within the downstream paired-end associated with the highest error rates in the control. Together, this suggests that many of these single occurrences are due to sequencing errors, rather than stochastic sampling of rare sequence variants.

These 184 patients were initially selected for this study since they were therapy naïve at the time their plasma sample was obtained, and because they eventually initiated HAART with a regimen containing only RT inhibitors. Our goal was to estimate the effect of undetected minority drug resistance on time-to-virologic failure, to determine if UDS had any added value over standard bulk sequence analysis with respect to treatment outcomes, and to enumerate a clinically relevant threshold for intra-patient resistant HIV-1 variants. To address these questions, we used a genotypic scoring method available through the Stanford HIV Drug Resistance Database to interpret drug resistance mutations detected by UDS and/or bulk sequence analysis [171]. Using the Stanford database scores, we estimated the number of active antiretrovirals within each RT inhibitor regimen based on the spectrum of resistance mutations observed within each patient sample, with <3 active antiretrovirals considered resistant. Using this scoring algorithm, only about 10% of patients received fewer than three active RT inhibitors at any given time over the course of follow-up. We found that these patients tended to experience virologic failure faster than patients that received fully-active HAART; however, results were not statistically significant (adjusted HR: 1.9, 95% CI: 0.90-4.1). Only 3 patients were

taking suboptimal HAART according to their bulk sequencing and UDS results, and 2 of 3 patients failed within the first 8 months of HAART. Patients receiving suboptimal HAART revealed by UDS alone also tended to experience virologic failure more quickly than patients without any resistance, but we were limited by the small number of patients in the exposed category (adjusted HR: 1.7, 95% CI: 0.75-4.0). We were also limited by the depth of sampling with respect to HIV-1 within individual patients. Since the resistance mutations we observed were either very high (>60%) or very low (<5%) abundance, we did not arrive at a clinical threshold for minority variants and virologic failure. We did, however, observe that very high copies of drug resistant virus (>10,000 co/mL) was associated with a decrease in time-to-virologic failure (adjusted HR: 3.5, 95% CI: 1.0-7.1).

6.2 IMPLICATIONS

In known control sequences, the large discrepancy in errors that we observed between analyses relying on raw sequences and Primer ID corrected consensus sequences highlighted the limitations of UDS technologies in their current form. Primer ID revealed that the number of HIV-1 RNA templates that we actually sampled is far lower than what we would have believed had we used raw sequence data alone. Thus, UDS is probably not able to reliably achieve the sampling depths that are often reported by others that rely on such data. These studies may filter errors from raw sequences based on control experiments and statistically defined cutoffs, but these methods do not account for differential amplification or PCR resampling. Measurement bias at all steps of UDS must be addressed before the effect of minority drug resistance on patient outcomes can be assessed.

However, it is also clear that Primer ID does not completely eliminate all errors, especially those that are associated with particular sequencing platforms.

This residual error is based on the composition of the nucleotide sequence itself, which is unknown for clinical samples, and its interaction with the specific UDS technology (ie., homopolymeric sequence regions). Residual, systematic errors associated with specific UDS platforms may skew estimates in unpredictable ways in large pools of unknown HIV-1 sequences. Further, for Primer ID corrected estimates to be accurate, we must make a number of assumptions. For example, excess Primer ID must be completely removed from each sample prior to amplification to prevent Primer ID sequences from being associated with multiple HIV-1 templates. If such a phenomenon were to occur, the viral population's sequence diversity could be deflated when consensus sequences are generated. Also, the Primer ID must be correctly sequenced so that a single Primer ID is not misread as multiple, unique Primer ID, thus inflating estimates of the size of the viral population.

When we used the Primer ID to more accurately measure RT inhibitor resistance in our population of 184 patients, we further showed how limited our ability is for detection of minority drug resistant variants using UDS. Untreated patients had relatively high viral loads compared to patients on therapy, which theoretically would allow the highest level of resolution of the HIV-1 population. Yet, we still only achieved the ability to reliably sample variants that were present among ≥7% of sequences on average. This is still superior to standard bulk sequence analysis (≥20%), and we identified a number of additional patients who were placed on suboptimal therapy that would not have been captured by standard sequencing methods. We did observe an increased risk of virologic failure among patients with minority variants that appeared to increase with copies of resistant virus or relative abundance, and these estimates were probably underestimates given their lack of statistical significance. However, to what degree these estimates would change

given a larger patient population or deeper sampling of the viral population is unknown.

6.3 Future Directions

Since this study's inception, the Primer ID has continued to improve. For example, the pre-assigned barcode sequence that differentiates patient samples is now represented twice within the cDNA primer, preventing incorrectly sequenced patient barcodes from either being discarded as non-matching, or worse, included with other similar patient barcode sequences as part of another sample. The Primer ID itself has been increased in length from 8 nucleotides to 11 nucleotides, which could facilitate increased sampling depth given the increase in available Primer ID sequence combinations from >65,000 to >4,000,000.

However, it is imperative that this method continue to be rigorously validated before it is expanded into additional epidemiologic or clinical use. Because Primer ID addresses sequencing errors from the PCR step forward, it is still unknown if the HIV-1 RNA is actually the same sequence. Known, HIV-1 RNA sequences that originate from viral culture and that replicate with complete fidelity must be generated to estimate the amount of error introduced during the cDNA synthesis step. Within the PCR step, we are still limited in terms of interpreting resistance mutations that are linked to the same HIV-1 genome because of the potential for PCR recombination. A mixture of control sequences with recombination markers include within every UDS run may allow us to at least determine the extent of this phenomenon. Finally, additional experiments are needed to determine the limit of our capacity to detect rare HIV-1 sequence variants given a set number of available templates. This is crucial to interpreting rare HIV-1 sequence variants in the context

of lower patient viral loads, such as might be expected for patients experiencing virologic failure.

Upon addressing these technical issues, Primer ID can and should be expanded to a larger number of patients, both prior to therapy and at treatment failure, to better estimate the point at which minority drug resistant variants begin to negatively affect patients receiving corresponding drugs. Clinically relevant thresholds are likely different not only between antiretroviral classes, but also for different individual resistance mutations and their combinations. This variability presents a major challenge to attempts at fleshing out meaningful estimates that clinical care providers can use to guide treatment selection. The Primer ID should also be expanded to other regions of the HIV-1 genome outside of RT that are drug targets. The very first study using Primer ID interrogated HIV-1 protease [26], but HIV-1 integrase and envelop could provide important information for clinicians considering regimens containing classes of drugs that target these enzymes. Although UDS is still far from supplanting current sequencing technology, Primer ID brings us a step closer to the level of accuracy needed for critical decision making.

APPENDIX 2.1 MUTATIONS ASSOCIATED WITH REVERSE TRANSCRIPTASE INHIBITOR RESISTANCE, INTERANTIONAL AIDS SOCIETY-USA, 2013 [183]

HIV-1 RT amino acid change[*]	NRTI								
							Multi-NRTI[†]		
	3TC FTC	ABC	AZT	d4T	ddI	TDF	+T69ins	+Q151M	TAM
M41L			X	X			X		X
A62V							X	X	
K65R	X	X		X	X	X			
D67N			X	X					X
K70R			X	X			X		X
K70E						X			
L74V		X			X				
V75I								X	
F77L								X	
Y115F		X							
F116Y								X	
M184V	X	X							
M184I	X								
L210W			X	X			X		X
T215Y/F			X	X			X		X
T219Q/E			X	X			X		X

	NNRTI			
	EFV	ETR	NVP	RPV
V90I		X		
A98G		X		
L100I	X	X	X	
K101E		X		X
K101P	X	X	X	X
K101H		X		
K103N/S	X		X	
V106A			X	
V106M	X	X	X	
V106I		X		
V108I	X		X	
E138A/G/K/Q		X		X
E138R				X
V179D/F/T		X		
V179L				X
Y181C/I	X	X	X	X
Y181V		X		X
Y188C/H			X	
Y188L	X		X	X
G190S	X	X		
G190A	X	X	X	
H221Y				X
P225H	X			
F227C				X

	NNRTI			
	EFV	ETR	NVP	RPV
M230I				X
M230L	X	X	X	X

RT=reverse transcriptase; NRTI=nucleoside reverse transcriptase inhibitor; NNRTI=non-NRTI; 3TC=lamivudine; FTC=emtricitabine; ABC=abacavir; AZT=zidovudine; d4T=stavudine; ddI=didanosine; TDF=tenofovir; ins=insertion; TAM=thymidine analogue mutation; EFV=efavirenz; ETR=etravirine; NVP=nevirapine; RPV=rilpivirine. *Amino acid changes are given in the following format: wild type amino acid, HIV-1 RT codon, mutant amino acid change(s). †Multi-NRTI resistance mutations confer resistance to all currently approved NRTI.

APPENDIX 2.2 SURVEILLANCE DRUG RESISTANCE MUTATIONS FOR EVIDENCE OF TRANSMITTED RESISTANCE, 2009 [189]

Amino acid change*	NRTI					
	3TC/FTC	ABC	AZT	d4T	ddI	TDF
M41L			X	X		
K65R	X	X		X	X	X
D67N/G/E			X	X		
T69ins/D	X	X	X	X	X	X
K70R			X	X		
K70E						X
L74V/I		X			X	
V75M/T/A/S				X	X	
F77L						
Y115F		X				
F116Y						
Q151M	X	X	X	X	X	X
M184V	X	X				
M184I	X					
L210W			X	X		
T215Y/F			X	X		
T215I/S/C/D/E/V*						
T219Q/E/N/R			X	X		

	NNRTI			
	EFV	ETR	NVP	RPV
L100I	X	X	X	
K101E		X		X
K101P	X	X	X	X
K103N/S	X		X	
V106A			X	
V106M	X	X	X	
V179F		X		
Y181C/I	X	X	X	X
Y181V		X		X
Y188C/H			X	
Y188L	X		X	X
G190S	X	X		
G190A	X	X	X	
G190E	X	X	X	X
H221Y				X
P225H	X			
M230L	X	X	X	X

NRTI=nucleoside(tide) reverse transcriptase inhibitor; NNRTI=non-NRTI; 3TC=lamivudine; FTC=emtricitabine; ABC=abacavir; AZT=zidovudine; d4T=stavudine; ddI=didanosine; TDF=tenofovir; ins=insertion; EFV=efavirenz; ETR=etravirine; NVP=nevirapine; RPV=rilpivirine. *T215 revertant mutations [185].

APPENDIX 2.3 STANFORD HIV DRUG RESISTANCE DATABASE ANTIRETROVIRAL SUSCEPTIBILITY SCORES FOR HIV-1 REVERSE TRANSCRIPTASE MUTATIONS [171]

	NRTI Score*					
Amino acid change†	3TC/FTC	ABC	AZT	d4T	ddI	TDF
E40F	0	5	5	5	5	5
M41L	5	5	15	15	5	5
E44A/D	5	5	5	5	5	5
A62V	5	5	5	5	5	5
K65R	30	45	-10	30	45	45
K65N	15	15	9	15	15	15
K66ins	30	45	45	45	45	45
K66del	15	15	15	15	15	15
D67N/G/E	0	5	15	15	5	5
D67S/T/H	0	5	10	10	5	5
D67ins	30	45	45	45	45	45
D67del	15	15	15	15	15	15
S68ins	30	45	45	45	45	45
S68del	15	15	15	15	15	15
T69D	0	0	0	10	30	0
T69N	0	0	5	5	10	0
T69G	0	10	10	10	10	10
T69ins	30	45	45	45	45	45
T69del	15	15	15	15	15	15
K70R	0	0	30	15	0	10
K70E/G	10	15	-10	0	15	25
K70T/S/N/Q	10	10	0	10	10	10
K70ins	30	45	45	45	45	45
K70del	15	15	15	15	15	15
W71ins	30	45	45	45	45	45
W71del	15	15	15	15	15	15
L74I	0	20	0	0	60	0
L74V	0	30	-10	0	60	0
V75A	0	0	0	15	15	0
V75I	5	0	0	10	10	0
V75M/T	0	0	0	60	30	0
V75S	0	0	0	20	10	0
F77L	5	10	10	10	10	5
Y115F	0	45	0	0	0	15
F116Y	5	10	10	10	10	5
V118I	0	5	5	5	5	5
Q151L	10	30	30	30	30	10
Q151M	15	60	60	60	60	15
M184I/V	60	15	-10	-10	10	-10

Amino acid change†	3TC/FTC	ABC	AZT	d4T	ddI	TDF
			NRTI Score*			
L210W	5	5	15	15	5	5
T215I/S/C/D/E/V/S	0	10	20	20	10	5
T215N/A/L	0	5	20	20	5	5
T215Y/F	5	15	45	45	15	15
T219Q/E/N/R/W/D/H	0	5	10	10	5	5
Combinations‡						
M41L+L210W	0	10	10	10	10	10
M41L+T215Y/F	0	10	10	10	10	10
M41L+T215X§	0	5	5	5	5	5
K65R+Y115F	0	0	0	0	0	10
K65R+Q151M	0	0	0	0	0	10
L74V/I+M184I/V	0	20	0	0	0	0
L210W+T215Y/F	0	10	10	10	10	10
L210W+T215X§	0	5	5	5	5	5

	NNRTI Score*			
	EFV	ETR	NVP	RPV
V90I	0	5	0	5
A98G	5	5	15	5
L100I	30	15	30	15
K101E	15	10	30	10
K101P	30	30	60	60
K101H	10	10	15	10
K103N/H	60	0	60	0
K103S	30	0	60	0
L103T	15	0	60	0
V106A	30	0	60	0
V106M	60	0	60	0
V108I	5	0	10	0
E138A	5	5	5	5
E138K	10	10	10	30
E138Q/G/R	5	10	5	15
V179D	10	10	10	10
V179E	10	5	10	5
V179T	0	5	0	5
V179F	5	15	10	15
Y181C	30	30	60	30
Y181I/V	30	60	60	60
Y188S	15	15	30	15
Y188C/H	15	0	60	0
Y188L	60	10	60	60
G190A	40	10	60	10
G190S/C/V/T	60	10	60	10
G190E/Q	60	15	60	15
H221Y	5	5	5	5
P225H	30	0	15	0

Amino acid change†	NNRTI Score* EFV	ETR	NVP	RPV
Combinations‡				
K103R+V179D	5	0	20	0
V106A+F227L	15	0	0	0
V106I+V179D	5	0	5	0
V179F+Y181C/I/V	0	15	0	15
Y181C/I/V+G190X§	0	15	0	15

NRTI=nucleoside(tide) reverse transcriptase inhibitor; NNRTI=non-NRTI; 3TC=lamivudine; FTC=emtricitabine; ABC=abacavir; AZT=zidovudine; d4T=stavudine; ddI=didanosine; TDF=tenofovir; ins=insertion; EFV=efavirenz; ETR=etravirine; NVP=nevirapine; RPV=rilpivirine. *To determine susceptibility to an individual antiretroviral, scores are added for each resistance mutation. A higher score corresponds to increased resistance to that antiretroviral. †Amino acid changes from are given in the following format: wild type amino acid, codon, resistant amino acid. ‡When combinations of given mutations are present, susceptibility scores for some antiretrovirals are increased over the score for the individual mutations. §For L215X, X is I/S/C/D/E/V/S/N/A/L; for G190X, X is A/S/E/Q/C/V/T.

APPENDIX 3.1 PRIMERS FOR cDNA SYNTHESIS AND DNA AMPLIFICATION

Application	Direction	HXB2*	Sequence (5'→3')
Patient samples			
cDNA†	antisense	2965-2992	GCCTTGCCAGCACGCTCA**CAGCTG**GCA-BBBB-CGNNNNNNNNTC-ACATTGTACTGATATCTAATYCCTGGTG
cDNA†	antisense	3258-3284	GCCTTGCCAGCACGCTCA**CAGCTG**GCA-BBBB-CGNNNNNNNNTC-CACTATAGGCTGTACTGTCCATTTATC
PCR 1/2	sense	2620-2647	GGCCATTGACAGAAGAAAAAATAAAAGC
PCR 1/2	sense	2965-2992	CACCAGGRATTAGATATCAGTACAATGT
PCR 1	antisense		GCCTTGCCAGCACGCTCACAG
PCR 2	antisense		CCAGCACGCTCACAGCTGGCA
Plasmid samples			
PCR 1	sense	2571-2598	GTACCAGTAAAATTAAAGCCAGGAATGG
PCR 1	antisense		GCCTTGCCAGCACGCTCAGGC
PCR 2	sense	2992-3284	CACCAGGGATTAGATATCAATATAATGT
PCR 2	antisense		CCAGCACGCTCAGGCCTTGCA

***HIV-1-specific sequence targeting each region were derived from primers used in Simen et al. [346]. †Primers used for cDNA sequence contained a *PvuII* recognition site (bold), a four-nucleotide sample barcode linked to each individual patient (B), a random 8-nucleotide Primer ID sequence (N), and 4 known nucleotides to facilitate alignment (underlined).**

APPENDIX 4.1 ERROR RATES ESTIMATED FROM DEEP SEQUENCING CONTROLS

	Raw sequence analysis					Primer ID consensus sequence analysis				
	Sequences	Errors	Nt*	Rate†	(95% CI)	Sequences	Errors	Nt*	Rate†	(95% CI)
454 Junior										
3,000	29,055	65,765	9,210,435	71	(71-72)	80	10	25,360	3.9	(1.9-7.3)
10,000	26,561	60,289	8,419,837	72	(71-72)	186	25	58,962	4.2	(2.7-6.3)
30,000	25,808	57,310	8,181,136	70	(69-71)	664	47	210,488	2.2	(1.6-3.0)
100,000	30,684	68,960	9,726,828	71	(70-71)	1,863	146	590,571	2.5	(2.1-2.9)
Total‡	112,108	252,324	35,538,236	71	(70-72)	2,893	228	885,381	2.6	(2.2-3.2)
MiSeq										
3,000, F	33,305	20,489	7,360,405	28	(27-28)	79	10	17,459	5.7	(2.7-11)
3,000, R	33,305	5,491	6,094,815	9.0	(8.8-9.3)	79	18	14,457	12	(7.4-20)
10,000, F	32,186	22,711	7,113,106	32	(32-32)	168	2	37,128	0.54	(0.0065-1.9)
10,000, R	32,186	4,274	5,890,038	7.3	(7.0-7.5)	168	9	30,744	2.9	(1.3-5.6)
30,000, F	28,318	23,162	6,258,278	37	(37-37)	656	6	144,976	0.41	(0.15-0.90)
30,000, R	28,318	3,635	5,182,194	7.0	(6.8-7.2)	656	30	120,048	2.5	(1.7-3.6)
100,000, F	30,013	103,490	6,632,873	56	(55-57)	1,807	8	399,347	0.20	(0.0086-0.40)
100,000, R	30,013	17,493	5,492,379	7.5	(7.2-7.7)	1,807	47	330,681	1.4	(1.0-1.9)
Total‡	123,822	120,983	50,024,088	24	(18-32)	2,710	130	1,094,840	1.2	(0.59-2.4)

***For the 454 Junior platform, the read length was 317 nucleotides (nt) spanning HIV-1 reverse transcriptase (RT) codons 34-139 (HXB2 nt 2648-2964). For the MiSeq platform, read lengths were 221 nt spanning HIV-1 RT codons 34-97 in the forward (F) direction (HXB2 nt 2648-2840), and 183 nt spanning codons 78-139 in the reverse (R) direction (HXB2 nt 2782-2964). †Rates are expressed as errors per 10,000 nt. ‡Standard errors for total error rates were calculated across all input DNA copy number dilutions using clustered sandwich estimators [351] .**

APPENDIX 5.1 ADJUSTED HAZARD RATIOS USING STRICTER INTERPRETATION OF DEEP SEQUENCING DATA

				Adjusted estimates			
Resistance to early therapy*	n	VF	p-m	HR†	95% CI	CLR	*P-value*
Predicted resistance							
<3 inactive RTI, Bulk or UDS	11	4	35	1.2	0.43-3.5	8.1	0.7
≥3 active RTI	170	40	681				
Predicted resistance							
<3 inactive RTI, Bulk & UDS	3	2	14	2.7	0.61-12	19	0.4
<3 inactive RTI, UDS only	8	2	21	0.77	0.18-3.3	18	
≥3 active RTI	170	40	681	1.0			
% Viral population, UDS							
≥20% resistant	6	3	22	2.1	0.63-7.4	12	0.4
<20% resistant	5	1	13	0.51	0.067-3.9	58	
None	171	41	697	1.0			
Copies resistant virus, UDS							
≥2,000 co/mL	9	3	22	1.2	0.35-3.8	11	0.8
<2,000 co/mL	173	42	710	1.0			
Copies resistant virus, UDS							
≥10,000 co/mL	6	3	22	2.2	0.63.-7.4	12	0.2
<10,000 co/mL	176	42	710	1.0			

VF=virologic failure; p-m=person-months; CLR=confidence interval ratio; RTI=reverse transcriptase inhibitor; UDS=ultra deep sequencing; MSM=men who have sex with men. *For deep sequencing, resistance mutations were considered genuine if they occurred >1 time within a patient sample. The number of active RTI per regimen was predicted using the Stanford HIV drug resistance database to interpret to interpret resistance mutations detected by bulk PCR sequence analysis and/or UDS [171]. †HR and 95% CI were adjusted for the following: (1) time-varying MSM and white race; (2) gender; and (3) age at therapy initiation modeled using restricted cubic splines with knots at 25th, 50th, and 75th percentiles.

REFERENCES

1. Egger M, May M, Chene G, Phillips AN, Ledergerber B, Dabis F*, et al.* Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *Lancet* 2002,**360**:119-129.

2. Mocroft A, Ledergerber B, Katlama C, Kirk O, Reiss P, d'Arminio Monforte A*, et al.* Decline in the AIDS and death rates in the EuroSIDA study: an observational study. *Lancet* 2003,**362**:22-29.

3. Hirsch MS, Brun-Vezinet F, Clotet B, Conway B, Kuritzkes DR, D'Aquila RT*, et al.* Antiretroviral drug resistance testing in adults infected with human immunodeficiency virus type 1: 2003 recommendations of an International AIDS Society-USA Panel. *Clin Infect Dis* 2003,**37**:113-128.

4. Grant RM, Hecht FM, Warmerdam M, Liu L, Liegler T, Petropoulos CJ*, et al.* Time trends in primary HIV-1 drug resistance among recently infected persons. *JAMA* 2002,**288**:181-188.

5. Little SJ, Holte S, Routy JP, Daar ES, Markowitz M, Collier AC*, et al.* Antiretroviral-drug resistance among patients recently infected with HIV. *N Engl J Med* 2002,**347**:385-394.

6. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents. *U.S. Department of Health and Human Services* 2013:1-267. Available at: http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentsGL.pdf. Accessed: March 2013.

7. Gunthard 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.

8. Hance AJ, Lemiale V, Izopet J, Lecossier D, Joly V, Massip P*, et al.* Changes in human immunodeficiency virus type 1 populations after treatment interruption in patients failing antiretroviral therapy. *J Virol* 2001,**75**:6410-6417.

9. Devereux HL, Youle M, Johnson MA, Loveday C. Rapid decline in detectability of HIV-1 drug resistance mutations after stopping therapy. *AIDS* 1999,**13**:F123-127.

10. Cane PA. Stability of transmitted drug-resistant HIV-1 species. *Curr Opin Infect Dis* 2005,**18**:537-542.

11. Cane P, Chrystie I, Dunn D, Evans B, Geretti AM, Green H*, et al.* Time trends in primary resistance to HIV drugs in the United Kingdom: multicentre observational study. *BMJ* 2005,**331**:1368.

12. Descamps D, Chaix ML, Andre P, Brodard V, Cottalorda J, Deveau C*, et al.* French national sentinel survey of antiretroviral drug resistance in patients with HIV-1 primary infection and in antiretroviral-naive chronically infected patients in 2001-2002. *J Acquir Immune Defic Syndr* 2005,**38**:545-552.

13. Barbour JD, Hecht FM, Wrin T, Liegler TJ, Ramstead CA, Busch MP*, et al.* Persistence of primary drug resistance among recently HIV-1 infected adults. *AIDS* 2004,**18**:1683-1689.

14. Brenner B, Routy JP, Quan Y, Moisi D, Oliveira M, Turner D*, et al.* Persistence of multidrug-resistant HIV-1 in primary infection leading to superinfection. *AIDS* 2004,**18**:1653-1660.

15. Delaugerre C, Morand-Joubert L, Chaix ML, Picard O, Marcelin AG, Schneider V*, et al.* Persistence of multidrug-resistant HIV-1 without antiretroviral treatment 2 years after sexual transmission. *Antivir Ther* 2004,**9**:415-421.

16. Ghosn J, Pellegrin I, Goujard C, Deveau C, Viard JP, Galimand J*, et al.* HIV-1 resistant strains acquired at the time of primary infection massively fuel the cellular reservoir and persist for lengthy periods of time. *AIDS* 2006,**20**:159-170.

17. Fox J, Dustan S, McClure M, Weber J, Fidler S. Transmitted drug-resistant HIV-1 in primary HIV-1 infection; incidence, evolution and impact on response to antiretroviral therapy. *HIV Med* 2006,**7**:477-483.

18. Li JZ, Paredes R, Ribaudo HJ, Svarovskaia ES, Metzner KJ, Kozal MJ*, et al.* Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis. *JAMA* 2011,**305**:1327-1335.

19. Liu SL, Rodrigo AG, Shankarappa R, Learn GH, Hsu L, Davidov O*, et al.* HIV quasispecies and resampling. *Science* 1996,**273**:415-416.

20. Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng* 2003,**96**:317-323.

21. Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* 1995,**267**:483-489.

22. Hoffmann C, Minkah N, Leipzig J, Wang G, Arens MQ, Tebas P*, et al.* DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Res* 2007,**35**:e91.

23. 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 Res* 2007,**17**:1195-1201.

24. Gilles A, Meglecz E, Pech N, Ferreira S, Malausa T, Martin JF. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* 2011,**12**:245.

25. Gianella S, Delport W, Pacold ME, Young JA, Choi JY, Little SJ*, et al.* Detection of minority resistance during early HIV-1 infection: natural variation and spurious detection rather than transmission and evolution of multiple viral variants. *J Virol* 2011,**85**:8359-8367.

26. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc Natl Acad Sci U S A* 2011,**108**:20166-20171.

27. Centers for Disease Control and Prevention. Pneumocystis pneumonia--Los Angeles. *MMWR* 1981,**30**:250-252.

28. UNAIDS. Report on the Global AIDS Epidemic, 2012. In: *Available at: http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/*. Geneva, Switzerland: World Health Oraganization; 2012. pp. 1-212.

29. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2011. In. Atlanta: U.S. Department of Health and Human Services; 2013. pp. 1-84.

30. Palella FJ, Jr., Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA*, et al.* Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med* 1998,**338**:853-860.

31. Centers for Disease Control and Prevention. AIDS Cases, Deaths, and Persons Living with AIDS,1985-2009, United States. In:

http://www.cdc.gov/HIV/topics/surveillance/resources/slides/trends/index.htm; 2009.

32. Centers for Disease Control and Prevention. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR Recomm Rep* 1992,**41**:1-19.

33. Centers for Disease Control and Prevention. Twenty-five years of HIV/AIDS--United States, 1981-2006. *MMWR* 2006,**55**:585-589.

34. Merriam-Webster Inc. *Merriam-Webster's collegiate dictionary*. 11th ed. Springfield, Mass.: Merriam-Webster, Inc.; 2003.

35. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2001. In. Atlanta: U.S. Department of Health and Human Services; 2002. pp. 1-50.

36. Centers for Disease Control and Prevention. Update: AIDS--United States, 2000. *MMWR* 2002,**51**:592-595.

37. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2002. In. Atlanta: U.S. Department of Health and Human Services; 2003. pp. 1-46.

38. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2003. In. Atlanta: U.S. Department of Health and Human Services; 2004. pp. 1-46.

39. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2004. In. Atlanta: U.S. Department of Health and Human Services; 2005. pp. 1-46.

40. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2005. In. Atlanta: U.S. Department of Health and Human Services; 2006. pp. 1-54.

41. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2006. In. Atlanta: U.S. Department of Health and Human Services; 2007. pp. 1-55.

42. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2007. In. Atlanta: U.S. Department of Health and Human Services; 2008. pp. 1-63.

43. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2008. In. Atlanta: U.S. Department of Health and Human Services; 2009. pp. 1-79.

44. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2009. In. Atlanta: U.S. Department of Health and Human Services; 2010. pp. 1-79.

45. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2010. In. Atlanta: U.S. Department of Health and Human Services; 2011. pp. 1-79.

46. Krawczyk CS, Funkhouser E, Kilby JM, Vermund SH. Delayed access to HIV diagnosis and care: Special concerns for the Southern United States. *AIDS Care* 2006,**18 Suppl 1**:S35-44.

47. Centers for Disease Control and Prevention. AIDS Public Information Data Set. In: U.S. Dept. of Health and Human Services; 2002.

48. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J*, et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983,**220**:868-871.

49. Fields BN, Knipe DM, Howley PM. *Fields' Virology*. 5th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2007.

50. Arien KK, Abraha A, Quinones-Mateu ME, Kestens L, Vanham G, Arts EJ. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. *J Virol* 2005,**79**:8979-8990.

51. Centers for Disease Control and Prevention. Human Immunodeficiency Virus Type 2. In; 1998.

52. Freed EO. HIV-1 replication. *Somat Cell Mol Genet* 2001,**26**:13-33.

53. Scarlatti G, Tresoldi E, Bjorndal A, Fredriksson R, Colognesi C, Deng HK*, et al.* In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. *Nat Med* 1997,**3**:1259-1265.

54. Adams M, Sharmeen L, Kimpton J, Romeo JM, Garcia JV, Peterlin BM*, et al.* Cellular latency in human immunodeficiency virus-infected individuals with high CD4 levels can be detected by the presence of promoter-proximal transcripts. *Proc Natl Acad Sci U S A* 1994,**91**:3862-3866.

55. Fauci AS, Pantaleo G, Stanley S, Weissman D. Immunopathogenic mechanisms of HIV infection. *Ann Intern Med* 1996,**124**:654-663.

56. Pantaleo G, Graziosi C, Fauci AS. New concepts in the immunopathogenesis of human immunodeficiency virus infection. *N Engl J Med* 1993,**328**:327-335.

57. Lyles RH, Munoz A, Yamashita TE, Bazmi H, Detels R, Rinaldo CR*, et al.* Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study. *J Infect Dis* 2000,**181**:872-880.

58. Pilcher CD, Joaki G, Hoffman IF, Martinson FE, Mapanje C, Stewart PW*, et al.* Amplified transmission of HIV-1: comparison of HIV-1 concentrations in semen and blood during acute and chronic infection. *AIDS* 2007,**21**:1723-1730.

59. Pilcher CD, Shugars DC, Fiscus SA, Miller WC, Menezes P, Giner J*, et al.* HIV in body fluids during primary HIV infection: implications for pathogenesis, treatment and public health. *AIDS* 2001,**15**:837-845.

60. Royce RA, Sena A, Cates W, Jr., Cohen MS. Sexual transmission of HIV. *N Engl J Med* 1997,**336**:1072-1078.

61. Vernazza PL, Troiani L, Flepp MJ, Cone RW, Schock J, Roth F*, et al.* Potent antiretroviral treatment of HIV-infection results in suppression of the seminal shedding of HIV. The Swiss HIV Cohort Study. *AIDS* 2000,**14**:117-121.

62. Porco TC, Martin JN, Page-Shafer KA, Cheng A, Charlebois E, Grant RM*, et al.* Decline in HIV infectivity following the introduction of highly active antiretroviral therapy. *AIDS* 2004,**18**:81-88.

63. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N*, et al.* Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med* 2011,**365**:493-505.

64. Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T*, et al.* The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat Med* 1996,**2**:1240-1243.

65. Wawer MJ, Makumbi F, Kigozi G, Serwadda D, Watya S, Nalugoda F*, et al.* Circumcision in HIV-infected men and its effect on HIV transmission to female partners in Rakai, Uganda: a randomised controlled trial. *Lancet* 2009,**374**:229-237.

66. MacDonald KS, Fowke KR, Kimani J, Dunand VA, Nagelkerke NJ, Ball TB*, et al.* Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. *J Infect Dis* 2000,**181**:1581-1589.

67. Fleming DT, Wasserheit JN. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. *Sex Transm Infect* 1999,**75**:3-17.

68. Schacker T, Little S, Connick E, Gebhard-Mitchell K, Zhang ZQ, Krieger J*, et al.* Rapid accumulation of human immunodeficiency virus (HIV) in lymphatic tissue reservoirs during acute and early HIV infection: implications for timing of antiretroviral therapy. *J Infect Dis* 2000,**181**:354-357.

69. Brenchley JM, Schacker TW, Ruff LE, Price DA, Taylor JH, Beilman GJ*, et al.* CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *J Exp Med* 2004,**200**:749-759.

70. Mehandru S, Poles MA, Tenner-Racz K, Horowitz A, Hurley A, Hogan C*, et al.* Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. *J Exp Med* 2004,**200**:761-770.

71. Grossman Z, Meier-Schellersheim M, Paul WE, Picker LJ. Pathogenesis of HIV infection: what the virus spares is as important as what it destroys. *Nat Med* 2006,**12**:289-295.

72. Little SJ, McLean AR, Spina CA, Richman DD, Havlir DV. Viral dynamics of acute HIV-1 infection. *J Exp Med* 1999,**190**:841-850.

73. Schacker T, Collier AC, Hughes J, Shea T, Corey L. Clinical and epidemiologic features of primary HIV infection. *Ann Intern Med* 1996,**125**:257-264.

74. Vergis EN, Mellors JW. Natural history of HIV-1 infection. *Infect Dis Clin North Am* 2000,**14**:809-+.

75. Pope M, Haase AT. Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. *Nat Med* 2003,**9**:847-852.

76. Blattner WA, Oursler KA, Cleghorn F, Charurat M, Sill A, Bartholomew C*, et al.* Rapid clearance of virus after acute HIV-1 infection: correlates of risk of AIDS. *J Infect Dis* 2004,**189**:1793-1801.

77. Sabin CA, Lundgren JD. The natural history of HIV infection. *Curr Opin HIV AIDS* 2013,**8**:311-317.

78. Collaborative Group on AIDS Incubation and HIV Survival. Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. *Lancet* 2000,**355**:1131-1137.

79. Biberfeld P, Porwit-Ksiazek A, Bottiger B, Morfeldt-Mansson L, Biberfeld G. Immunohistopathology of lymph nodes in HTLV-III infected homosexuals with persistent adenopathy or AIDS. *Cancer Res* 1985,**45**:4665s-4670s.

80. Longini IM, Jr., Clark WS, Gardner LI, Brundage JF. The dynamics of CD4+ T-lymphocyte decline in HIV-infected individuals: a Markov modeling approach. *J Acquir Immune Defic Syndr* 1991,**4**:1141-1147.

81. Osmond D, Charlebois E, Lang W, Shiboski S, Moss A. Changes in AIDS survival time in two San Francisco cohorts of homosexual men, 1983 to 1993. *JAMA* 1994,**271**:1083-1087.

82. Mocroft AJ, Lundgren JD, d'Armino Monforte A, Ledergerber B, Barton SE, Vella S*, et al.* Survival of AIDS patients according to type of AIDS-defining event. The AIDS in Europe Study Group. *Int J Epidemiol* 1997,**26**:400-407.

83. Ledergerber B, Egger M, Opravil M, Telenti A, Hirschel B, Battegay M*, et al.* Clinical progression and virological failure on highly active antiretroviral therapy in HIV-1 patients: a prospective cohort study. Swiss HIV Cohort Study. *Lancet* 1999,**353**:863-868.

84. Lima VD, Hogg RS, Harrigan PR, Moore D, Yip B, Wood E*, et al.* Continued improvement in survival among HIV-infected individuals with newer forms of highly active antiretroviral therapy. *AIDS* 2007,**21**:685-692.

85. Murphy EL, Collier AC, Kalish LA, Assmann SF, Para MF, Flanigan TP*, et al.* Highly active antiretroviral therapy decreases mortality and morbidity in patients with advanced HIV disease. *Ann Intern Med* 2001,**135**:17-26.

86. U.S. Food and Drug Administration. http://www.fda.gov/oashi/AIDS/virals.html. In; 2008.

87. Hammer SM, Saag MS, Schechter M, Montaner JS, Schooley RT, Jacobsen DM*, et al.* Treatment for adult HIV infection: 2006 recommendations of the International AIDS Society-USA panel. *JAMA* 2006,**296**:827-843.

88. Riddler SA, Haubrich R, DiRienzo G, Peeples L, Powderly WG, Klingman KL*, et al.* A prospective, randomized, Phase III trial of NRTI-, PI-, and NNRTI-sparing regimens for initial treatment of HIV-1 infection - ACTG 5142. In: *Presented at: XVI International AIDS Conference*. Aug 13-18, 2006; Toronto, Canada. Abstract THLB0204.

89. Squires KE. An introduction to nucleoside and nucleotide analogues. *Antivir Ther* 2001,**6 Suppl 3**:1-14.

90. De Clercq E. New approaches toward anti-HIV chemotherapy. *J Med Chem* 2005,**48**:1297-1313.

91. Balzarini J. Current status of the non-nucleoside reverse transcriptase inhibitors of human immunodeficiency virus type 1. *Curr Top Med Chem* 2004,**4**:921-944.

92. Gazzard B, Bernard AJ, Boffito M, Churchill D, Edwards S, Fisher N*, et al.* British HIV Association (BHIVA) guidelines for the treatment of HIV-infected adults with antiretroviral therapy (2006). *HIV Med* 2006,**7**:487-503.

93. Severe P, Juste MA, Ambroise A, Eliacin L, Marchand C, Apollon S*, et al.* Early versus standard antiretroviral therapy for HIV-infected adults in Haiti. *N Engl J Med* 2010,**363**:257-265.

94. Gay CL, Napravnik S, Eron JJ, Jr. Advanced immunosuppression at entry to HIV care in the southeastern United States and associated risk factors. *AIDS* 2006,**20**:775-778.

95. Perelson AS, Essunger P, Cao Y, Vesanen M, Hurley A, Saksela K*, et al.* Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* 1997,**387**:188-191.

96. 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.

97. Blankson JN, Finzi D, Pierson TC, Sabundayo BP, Chadwick K, Margolick JB*, et al.* Biphasic decay of latently infected CD4+ T cells in acute human immunodeficiency virus type 1 infection. *J Infect Dis* 2000,**182**:1636-1642.

98. Greub G, Cozzi-Lepri A, Ledergerber B, Staszewski S, Perrin L, Miller V*, et al.* Intermittent and sustained low-level HIV viral rebound in patients receiving potent antiretroviral therapy. *AIDS* 2002,**16**:1967-1969.

99. Markowitz M, Louie M, Hurley A, Sun E, Di Mascio M, Perelson AS*, et al.* A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo. *J Virol* 2003,**77**:5037-5038.

100. Murray JM, Emery S, Kelleher AD, Law M, Chen J, Hazuda DJ*, et al.* Antiretroviral therapy with the integrase inhibitor raltegravir alters decay kinetics of HIV, significantly reducing the second phase. *AIDS* 2007,**21**:2315-2321.

101. Wu H, Kuritzkes DR, McClernon DR, Kessler H, Connick E, Landay A*, et al.* Characterization of viral dynamics in human immunodeficiency virus type 1-infected patients treated with combination antiretroviral therapy: relationships to host factors, cellular restoration, and virologic end points. *J Infect Dis* 1999,**179**:799-807.

102. Palmisano L, Giuliano M, Nicastri E, Pirillo MF, Andreotti M, Galluzzo CM*, et al.* Residual viraemia in subjects with chronic HIV infection and viral load < 50 copies/ml: the impact of highly active antiretroviral therapy. *AIDS* 2005,**19**:1843-1847.

103. Zhang L, Ramratnam B, Tenner-Racz K, He Y, Vesanen M, Lewin S*, et al.* Quantifying residual HIV-1 replication in patients receiving combination antiretroviral therapy. *N Engl J Med* 1999,**340**:1605-1613.

104. Di Mascio M, Markowitz M, Louie M, Hogan C, Hurley A, Chung C*, et al.* Viral blip dynamics during highly active antiretroviral therapy. *J Virol* 2003,**77**:12165-12172.

105. Metzner KJ. Persistence of drug-resistant HIV-1 and possible implications for antiretroviral therapy. *Future Virology* 2006,**1**:377-391.

106. Li JZ, Gallien S, Do TD, Martin JN, Deeks S, Kuritzkes DR*, et al.* Prevalence and significance of HIV-1 drug resistance mutations among patients on antiretroviral therapy with detectable low-level viremia. *Antimicrob Agents Chemother* 2012,**56**:5998-6000.

107. Easterbrook PJ, Ives N, Waters A, Mullen J, O'Shea S, Peters B*, et al.* The natural history and clinical significance of intermittent viraemia in patients with initial viral suppression to < 400 copies/ml. *AIDS* 2002,**16**:1521-1527.

108. Havlir DV, Bassett R, Levitan D, Gilbert P, Tebas P, Collier AC*, et al.* Prevalence and predictive value of intermittent viremia with combination HIV therapy. *JAMA* 2001,**286**:171-179.

109. Sklar PA, Ward DJ, Baker RK, Wood KC, Gafoor Z, Alzola CF*, et al.* Prevalence and clinical correlates of HIV viremia ('blips') in patients with previous suppression below the limits of quantification. *AIDS* 2002,**16**:2035-2041.

110. Mocroft A, Youle M, Moore A, Sabin CA, Madge S, Lepri AC*, et al.* Reasons for modification and discontinuation of antiretrovirals: results from a single treatment centre. *AIDS* 2001,**15**:185-194.

111. Kempf MC, Pisu M, Dumcheva A, Westfall AO, Kilby JM, Saag MS. Gender differences in discontinuation of antiretroviral treatment regimens. *J Acquir Immune Defic Syndr* 2009,**52**:336-341.

112. Robison LS, Westfall AO, Mugavero MJ, Kempf MC, Cole SR, Allison JJ*, et al.* Short-term discontinuation of HAART regimens more common in vulnerable patient populations. *AIDS Res Hum Retroviruses* 2008,**24**:1347-1355.

113. Florence E, Lundgren J, Dreezen C, Fisher M, Kirk O, Blaxhult A*, et al.* Factors associated with a reduced CD4 lymphocyte count response to HAART despite full viral suppression in the EuroSIDA study. *HIV Med* 2003,**4**:255-262.

114. Deeks SG, Barbour JD, Martin JN, Swanson MS, Grant RM. Sustained CD4+ T cell response after virologic failure of protease inhibitor-based regimens in patients with human immunodeficiency virus infection. *J Infect Dis* 2000,**181**:946-953.

115. Nettles RE, Kieffer TL, Kwon P, Monie D, Han Y, Parsons T*, et al.* Intermittent HIV-1 viremia (Blips) and drug resistance in patients receiving HAART. *JAMA* 2005,**293**:817-829.

116. Clavel F, Hance AJ. HIV drug resistance. *N Engl J Med* 2004,**350**:1023-1035.

117. Nora T, Charpentier C, Tenaillon O, Hoede C, Clavel F, Hance AJ. Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. *J Virol* 2007,**81**:7620-7628.

118. Charpentier C, Nora T, Tenaillon O, Clavel F, Hance AJ. Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *J Virol* 2006,**80**:2472-2482.

119. 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.

120. Nijhuis M, Deeks S, Boucher C. Implications of antiretroviral resistance on viral fitness. *Curr Opin Infect Dis* 2001,**14**:23-28.

121. Quiñones-Mateu ME, Arts EJ. HIV-1 Fitness: Implications for Drug Resistance, Disease Progression, and Global Epidemic Evolution. In: *HIV Sequence Compendium*. Edited by Kuiken C, Foley B, Hahn B*, et al.* Los Alamos, NM: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory; 2001. pp. 134-170.

122. Deval J, White KL, Miller MD, Parkin NT, Courcambeck J, Halfon P*, et al.* Mechanistic basis for reduced viral and enzymatic fitness of HIV-1 reverse transcriptase containing both K65R and M184V mutations. *J Biol Chem* 2004,**279**:509-516.

123. Wei X, Liang C, Gotte M, Wainberg MA. Negative effect of the M184V mutation in HIV-1 reverse transcriptase on initiation of viral DNA synthesis. *Virology* 2003,**311**:202-212.

124. Arts EJ, Wainberg MA. Mechanisms of nucleoside analog antiviral activity and resistance during human immunodeficiency virus reverse transcription. *Antimicrob Agents Chemother* 1996,**40**:527-540.

125. Paintsil E, Margolis A, Collins JA, Alexander L. The contribution of HIV fitness to the evolution pattern of reverse transcriptase inhibitor resistance. *J Med Virol* 2006,**78**:425-430.

126. Hu Z, Giguel F, Hatano H, Reid P, Lu J, Kuritzkes DR. Fitness comparison of thymidine analog resistance pathways in human immunodeficiency virus type 1. *J Virol* 2006,**80**:7020-7027.

127. Palmer S, Boltz V, Martinson N, Maldarelli F, Gray G, McIntyre J*, et al.* Persistence of nevirapine-resistant HIV-1 in women after single-dose nevirapine therapy for prevention of maternal-to-fetal HIV-1 transmission. *Proc Natl Acad Sci U S A* 2006,**103**:7094-7099.

128. Little SJ, Frost SD, Wong JK, Smith DM, Pond SL, Ignacio CC*, et al.* Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection. *J Virol* 2008,**82**:5510-5518.

129. Martinez-Picado J, Martinez MA. HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: a view from the clinic and ex vivo. *Virus Res* 2008,**134**:104-123.

130. Rhee SY, Fessel WJ, Zolopa AR, Hurley L, Liu T, Taylor J*, et al.* HIV-1 Protease and reverse-transcriptase mutations: correlations with antiretroviral therapy in subtype B isolates and implications for drug-resistance surveillance. *J Infect Dis* 2005,**192**:456-465.

131. Larder BA, Darby G, Richman DD. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. *Science* 1989,**243**:1731-1734.

132. 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.

133. White KL, Margot NA, Ly JK, Chen JM, Ray AS, Pavelko M*, et al.* A combination of decreased NRTI incorporation and decreased excision determines the resistance profile of HIV-1 K65R RT. *AIDS* 2005,**19**:1751-1760.

134. White KL, Margot NA, Wrin T, Petropoulos CJ, Miller MD, Naeger LK. Molecular mechanisms of resistance to human immunodeficiency virus type 1 with reverse transcriptase mutations K65R and K65R+M184V and their effects on enzyme function and viral replication capacity. *Antimicrob Agents Chemother* 2002,**46**:3437-3446.

135. Johnson VA, Brun-Vezinet F, Clotet B, Gunthard HF, Kuritzkes DR, Pillay D*, et al.* Update of the Drug Resistance Mutations in HIV-1: Spring 2008. *Top HIV Med* 2008,**16**:62-68.

136. Lennerstrand J, Stammers DK, Larder BA. Biochemical mechanism of human immunodeficiency virus type 1 reverse transcriptase resistance to stavudine. *Antimicrob Agents Chemother* 2001,**45**:2144-2146.

137. Menendez-Arias L. Mechanisms of resistance to nucleoside analogue inhibitors of HIV-1 reverse transcriptase. *Virus Res* 2008,**134**:124-146.

138. Gotte M, Arion D, Parniak MA, Wainberg MA. The M184V mutation in the reverse transcriptase of human immunodeficiency virus type 1 impairs rescue of chain-terminated DNA synthesis. *J Virol* 2000,**74**:3579-3585.

139. Hsiou Y, Ding J, Das K, Clark AD, Jr., Hughes SH, Arnold E. Structure of unliganded HIV-1 reverse transcriptase at 2.7 A resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure* 1996,**4**:853-860.

140. Jacobo-Molina A, Ding J, Nanni RG, Clark AD, Jr., Lu X, Tantillo C*, et al.* Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc Natl Acad Sci U S A* 1993,**90**:6320-6324.

141. Domaoal RA, Demeter LM. Structural and biochemical effects of human immunodeficiency virus mutants resistant to non-nucleoside reverse transcriptase inhibitors. *Int J Biochem Cell Biol* 2004,**36**:1735-1751.

142. Hsiou Y, Ding J, Das K, Clark AD, Jr., Boyer PL, Lewi P*, et al.* The Lys103Asn mutation of HIV-1 RT: a novel mechanism of drug resistance. *J Mol Biol* 2001,**309**:437-445.

143. Ren J, Nichols C, Bird L, Chamberlain P, Weaver K, Short S*, et al.* Structural mechanisms of drug resistance for mutations at codons 181 and 188 in HIV-1 reverse transcriptase and the improved resilience of second generation non-nucleoside inhibitors. *J Mol Biol* 2001,**312**:795-805.

144. Hirsch MS, Gunthard HF, Schapiro JM, Brun-Vezinet F, Clotet B, Hammer SM*, et al.* Antiretroviral Drug Resistance Testing in Adult HIV-1 Infection: 2008 Recommendations of an International AIDS Society-USA Panel. *Clin Infect Dis* 2008.

145. Yerly S, Kaiser L, Race E, Bru JP, Clavel F, Perrin L. Transmission of antiretroviral-drug-resistant HIV-1 variants. *Lancet* 1999,**354**:729-733.

146. 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.

147. Harrigan PR, Wynhoven B, Brumme ZL, Brumme CJ, Sattha B, Major JC*, et al.* HIV-1 drug resistance: degree of underestimation by a cross-sectional versus a longitudinal testing approach. *J Infect Dis* 2005,**191**:1325-1330.

148. Richman DD, Morton SC, Wrin T, Hellmann N, Berry S, Shapiro MF*, et al.* The prevalence of antiretroviral drug resistance in the United States. *AIDS* 2004,**18**:1393-1401.

149. Cheung PK, Wynhoven B, Harrigan PR. 2004: which HIV-1 drug resistance mutations are common in clinical practice? *AIDS Rev* 2004,**6**:107-116.

150. de Mendoza C, Garrido C, Corral A, Ramirez-Olivencia G, Jimenez-Nacher I, Zahonero N*, et al.* Changing rates and patterns of drug resistance mutations in antiretroviral-experienced HIV-infected patients. *AIDS Res Hum Retroviruses* 2007,**23**:879-885.

151. Kagan R, Winters M, Merigan T, Heseltine P. HIV type 1 genotypic resistance in a clinical database correlates with antiretroviral utilization. *AIDS Res Hum Retroviruses* 2004,**20**:1-9.

152. Phillips AN, Dunn D, Sabin C, Pozniak A, Matthias R, Geretti AM*, et al.* Long term probability of detection of HIV-1 drug resistance after starting antiretroviral therapy in routine clinical practice. *AIDS* 2005,**19**:487-494.

153. Santoro MM, Svicher V, Gori C, Zaccarelli M, Tozzi V, Forbici F*, et al.* Temporal characterization of drug resistance associated mutations in HIV-1 protease and reverse transcriptase in patients failing antiretroviral therapy. *New Microbiol* 2006,**29**:89-100.

154. Tamalet C, Fantini J, Tourres C, Yahi N. Resistance of HIV-1 to multiple antiretroviral drugs in France: a 6-year survey (1997-2002) based on an analysis of over 7000 genotypes. *AIDS* 2003,**17**:2383-2388.

155. Napravnik S, Keys JR, Quinlivan EB, Wohl DA, Mikeal OV, Eron JJ, Jr. Triple-class antiretroviral drug resistance: risk and predictors among HIV-1-infected patients. *AIDS* 2007,**21**:825-834.

156. Zaccarelli M, Tozzi V, Lorenzini P, Trotta MP, Forbici F, Visco-Comandini U*, et al.* Multiple drug class-wide resistance associated with poorer survival after treatment failure in a cohort of HIV-infected patients. *AIDS* 2005,**19**:1081-1089.

157. Costagliola D, Descamps D, Assoumou L, Morand-Joubert L, Marcelin AG, Brodard V*, et al.* Prevalence of HIV-1 drug resistance in treated patients: a French nationwide study. *J Acquir Immune Defic Syndr* 2007,**46**:12-18.

158. Gange SJ, Schneider MF, Grant RM, Liegler T, French A, Young M*, et al.* Genotypic resistance and immunologic outcomes among HIV-1-infected women with viral failure. *J Acquir Immune Defic Syndr* 2006,**41**:68-74.

159. Gallego O, de Mendoza C, Corral A, Soriano V. Prevalence of drug resistance genotypes causing broad cross-resistance to nucleos(t)ide analogues. *AIDS* 2004,**18**:689-690.

160. Camacho R, Theys K, Abecasis A, Deforche K, Carvalho AP, Cabanas J*, et al.* Rise and fall of the RT K65R incidence in the Portugese resistance database [Abstract 121]. *Antivir Ther* 2006,**11**:S134.

161. de Mendoza C, Jimenez-Nacher I, Garrido C, Barreiro P, Poveda E, Corral A*, et al.* Changing patterns in HIV reverse transcriptase resistance mutations after availability of tenofovir. *Clin Infect Dis* 2008,**46**:1782-1785.

162. Delaugerre C, Rohban R, Simon A, Mouroux M, Tricot C, Agher R*, et al.* Resistance profile and cross-resistance of HIV-1 among patients failing a non-nucleoside reverse transcriptase inhibitor-containing regimen. *J Med Virol* 2001,**65**:445-448.

163. Kantor R, Shafer RW, Follansbee S, Taylor J, Shilane D, Hurley L*, et al.* Evolution of resistance to drugs in HIV-1-infected patients failing antiretroviral therapy. *AIDS* 2004,**18**:1503-1511.

164. Napravnik S, Edwards D, Stewart P, Stalzer B, Matteson E, Eron JJ, Jr. HIV-1 drug resistance evolution among patients on potent combination antiretroviral therapy with detectable viremia. *J Acquir Immune Defic Syndr* 2005,**40**:34-40.

165. Cozzi-Lepri A, Phillips AN, Ruiz L, Clotet B, Loveday C, Kjaer J*, et al.* Evolution of drug resistance in HIV-infected patients remaining on a virologically failing combination antiretroviral therapy regimen. *AIDS* 2007,**21**:721-732.

166. Hatano H, Hunt P, Weidler J, Coakley E, Hoh R, Liegler T*, et al.* Rate of viral evolution and risk of losing future drug options in heavily pretreated, HIV-infected patients who continue to receive a stable, partially suppressive treatment regimen. *Clin Infect Dis* 2006,**43**:1329-1336.

167. Metzner KJ, Allers K, Rauch P, Harrer T. Rapid selection of drug-resistant HIV-1 during the first months of suppressive ART in treatment-naive patients. *AIDS* 2007,**21**:703-711.

168. Grabar S, Le Moing V, Goujard C, Leport C, Kazatchkine MD, Costagliola D*, et al.* Clinical outcome of patients with HIV-1 infection according to immunologic and virologic response after 6 months of highly active antiretroviral therapy. *Ann Intern Med* 2000,**133**:401-410.

169. Haubrich RH, Kemper CA, Hellmann NS, Keiser PH, Witt MD, Forthal DN*, et al.* The clinical relevance of non-nucleoside reverse transcriptase inhibitor hypersusceptibility: a prospective cohort analysis. *AIDS* 2002,**16**:F33-40.

170. Lucas GM, Gallant JE, Moore RD. Relationship between drug resistance and HIV-1 disease progression or death in patients undergoing resistance testing. *AIDS* 2004,**18**:1539-1548.

171. Liu TF, Shafer RW. Web resources for HIV type 1 genotypic-resistance test interpretation. *Clin Infect Dis* 2006,**42**:1608-1618.

172. Johnson VA, Brun-Vezinet F, Clotet B, Kuritzkes DR, Pillay D, Schapiro JM*, et al.* Update of the drug resistance mutations in HIV-1: Fall 2006. *Top HIV Med* 2006,**14**:125-130.

173. Bracciale L, Di Giambenedetto S, Colafigli M, La Torre G, Prosperi M, Santangelo R*, et al.* Virological suppression reduces clinical progression in patients with multiclass-resistant HIV type 1. *AIDS Res Hum Retroviruses* 2009,**25**:261-267.

174. Castor D, Vlahov D, Hoover DR, Berkman A, Wu YF, Zeller B*, et al.* The relationship between genotypic sensitivity score and treatment outcomes in late stage HIV disease after supervised HAART. *J Med Virol* 2009,**81**:1323-1335.

175. Hogg RS, Bangsberg DR, Lima VD, Alexander C, Bonner S, Yip B*, et al.* Emergence of drug resistance is associated with an increased risk of death among patients first starting HAART. *PLoS Med* 2006,**3**:e356.

176. Kozal MJ, Hullsiek KH, Macarthur RD, Berg-Wolf M, Peng G, Xiang Y*, et al.* The Incidence of HIV drug resistance and its impact on progression of HIV disease among antiretroviral-naive participants started on three different antiretroviral therapy strategies. *HIV Clin Trials* 2007,**8**:357-370.

177. Hecht FM, Grant RM, Petropoulos CJ, Dillon B, Chesney MA, Tian H*, et al.* Sexual transmission of an HIV-1 variant resistant to multiple reverse-transcriptase and protease inhibitors. *N Engl J Med* 1998,**339**:307-311.

178. Colgrove RC, Pitt J, Chung PH, Welles SL, Japour AJ. Selective vertical transmission of HIV-1 antiretroviral resistance mutations. *AIDS* 1998,**12**:2281-2288.

179. Veenstra J, Schuurman R, Cornelissen M, van't Wout AB, Boucher CA, Schuitemaker H*, et al.* Transmission of zidovudine-resistant human immunodeficiency virus type 1 variants following deliberate injection of blood from a patient with AIDS: characteristics and natural history of the virus. *Clin Infect Dis* 1995,**21**:556-560.

180. Yerly S, von Wyl V, Ledergerber B, Boni J, Schupbach J, Burgisser P*, et al.* Transmission of HIV-1 drug resistance in Switzerland: a 10-year molecular epidemiology survey. *AIDS* 2007,**21**:2223-2229.

181. Shafer RW, Rhee SY, Bennett DE. Consensus drug resistance mutations for epidemiological surveillance: basic principles and potential controversies. *Antivir Ther* 2008,**13 Suppl 2**:59-68.

182. Geretti AM. Epidemiology of antiretroviral drug resistance in drug-naive persons. *Curr Opin Infect Dis* 2007,**20**:22-32.

183. Johnson VA, Calvez V, Gunthard HF, Paredes R, Pillay D, Shafer RW*, et al.* Update of the drug resistance mutations in HIV-1: March 2013. *Top Antivir Med* 2013,**21**:6-14.

184. Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 2003,**31**:298-303.

185. 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.

186. Yerly S, Rakik A, De Loes SK, Hirschel B, Descamps D, Brun-Vezinet F*, et al.* Switch to unusual amino acids at codon 215 of the human immunodeficiency virus type 1 reverse transcriptase gene in seroconvertors infected with zidovudine-resistant variants. *J Virol* 1998,**72**:3520-3523.

187. Mihailidis C, Dunn D, Pillay D, Pozniak A. Effect of isolated V118I mutation in reverse transcriptase on response to first-line antiretroviral therapy. *AIDS* 2008,**22**:427-430.

188. Shafer RW, Rhee SY, Pillay D, Miller V, Sandstrom P, Schapiro JM*, et al.* HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *AIDS* 2007,**21**:215-223.

189. Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Fleury H, Kiuchi M*, et al.* Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS ONE* 2009,**4**:e4724.

190. Fox JM, Fidler S, Weber J. Resistance to HIV drugs in UK may be lower in some areas. *BMJ* 2006,**332**:179-180.

191. Jayaraman GC, Archibald CP, Kim J, Rekart ML, Singh AE, Harmen S*, et al.* A population-based approach to determine the prevalence of transmitted drug-resistant HIV among recent versus established HIV infections: results from the Canadian HIV strain and drug resistance surveillance program. *J Acquir Immune Defic Syndr* 2006,**42**:86-90.

192. Novak RM, Chen L, MacArthur RD, Baxter JD, Huppler Hullsiek K, Peng G*, et al.* Prevalence of antiretroviral drug resistance mutations in chronically HIV-infected, treatment-naive patients: implications for routine resistance screening before initiation of antiretroviral therapy. *Clin Infect Dis* 2005,**40**:468-474.

193. SPREAD programme. Transmission of drug-resistant HIV-1 in Europe remains limited to single classes. *AIDS* 2008,**22**:625-635.

194. Weinstock HS, Zaidi I, Heneine W, Bennett D, Garcia-Lerma JG, Douglas JM, Jr.*, et al.* The epidemiology of antiretroviral drug resistance among drug-naive HIV-1-infected persons in 10 US cities. *J Infect Dis* 2004,**189**:2174-2180.

195. Booth CL, Geretti AM. Prevalence and determinants of transmitted antiretroviral drug resistance in HIV-1 infection. *J Antimicrob Chemother* 2007,**59**:1047-1056.

196. Toni T, Masquelier B, Minga A, Anglaret X, Danel C, Coulibaly A*, et al.* HIV-1 antiretroviral drug resistance in recently infected patients in Abidjan, Cote d'Ivoire: A 4-year survey, 2002-2006. *AIDS Res Hum Retroviruses* 2007,**23**:1155-1160.

197. Sigaloff KC, Mandaliya K, Hamers RL, Otieno F, Jao IM, Lyagoba F*, et al.* Short communication: High prevalence of transmitted antiretroviral drug resistance among newly HIV type 1 diagnosed adults in Mombasa, Kenya. *AIDS Res Hum Retroviruses* 2012,**28**:1033-1037.

198. Ssemwanga D, Kapaata A, Lyagoba F, Magambo B, Nanyonjo M, Mayanja BN*, et al.* Low drug resistance levels among drug-naive individuals with recent HIV type 1 infection in a rural clinical cohort in southwestern Uganda. *AIDS Res Hum Retroviruses* 2012,**28**:1784-1787.

199. Ndembi N, Hamers RL, Sigaloff KC, Lyagoba F, Magambo B, Nanteza B*, et al.* Transmitted antiretroviral drug resistance among newly HIV-1 diagnosed young individuals in Kampala. *AIDS* 2011,**25**:905-910.

200. Nazziwa J, Njai HF, Ndembi N, Birungi J, Lyagoba F, Gershim A*, et al.* Short communication: HIV type 1 transmitted drug resistance and evidence of transmission clusters among recently infected antiretroviral-naive individuals from Ugandan fishing communities of Lake Victoria. *AIDS Res Hum Retroviruses* 2013,**29**:788-795.

201. Price MA, Wallis CL, Lakhi S, Karita E, Kamali A, Anzala O*, et al.* Transmitted HIV type 1 drug resistance among individuals with recent HIV infection in East and Southern Africa. *AIDS Res Hum Retroviruses* 2011,**27**:5-12.

202. Zhang J, Kang D, Fu J, Sun X, Lin B, Bi Z*, et al.* Surveillance of transmitted HIV type 1 drug resistance in newly diagnosed hiv type 1-infected patients in Shandong Province, China. *AIDS Res Hum Retroviruses* 2010,**26**:99-103.

203. Apisarnthanarak A, Jirayasethpong T, Sa-nguansilp C, Thongprapai H, Kittihanukul C, Kamudamas A*, et al.* Antiretroviral drug resistance among antiretroviral-naive persons with recent HIV infection in Thailand. *HIV Med* 2008,**9**:322-325.

204. Masquelier B, Bhaskaran K, Pillay D, Gifford R, Balestre E, Jorgensen LB*, et al.* Prevalence of transmitted HIV-1 drug resistance and the role of resistance algorithms: data from seroconverters in the CASCADE collaboration from 1987 to 2003. *J Acquir Immune Defic Syndr* 2005,**40**:505-511.

205. Bartmeyer B, Kuecherer C, Houareau C, Werning J, Keeren K, Somogyi S*, et al.* Prevalence of transmitted drug resistance and impact of transmitted resistance on treatment success in the German HIV-1 Seroconverter Cohort. *PLoS ONE* 2010,**5**:e12718.

206. Colafigli M, Torti C, Trecarichi EM, Albini L, Rosi A, Micheli V*, et al.* Evolution of transmitted HIV-1 drug resistance in HIV-1-infected patients in Italy from 2000 to 2010. *Clin Microbiol Infect* 2012,**18**:E299-304.

207. Temereanca A, Ene L, Mehta S, Manolescu L, Duiculescu D, Ruta S. Transmitted HIV drug resistance in treatment-naive Romanian patients. *J Med Virol* 2013.

208. Lunar MM, Zidovec Lepej S, Abecasis AB, Tomazic J, Vidmar L, Karner P*, et al.* Short communication: prevalence of HIV type 1 transmitted drug resistance in Slovenia: 2005-2010. *AIDS Res Hum Retroviruses* 2013,**29**:343-349.

209. de Mendoza C, Rodriguez C, Colomina J, Tuset C, Garcia F, Eiros JM*, et al.* Resistance to nonnucleoside reverse-transcriptase inhibitors and prevalence of HIV type 1 non-B subtypes are increasing among persons with recent infection in Spain. *Clin Infect Dis* 2005,**41**:1350-1354.

210. UK Collaborative Group on HIV Drug Resistance, UK Collaborative HIV Cohort Study, UK Register of HIV Seroconverters. Evidence of a decline in transmitted HIV-1 drug resistance in the United Kingdom. *AIDS* 2007,**21**:1035-1039.

211. Booth CL, Garcia-Diaz AM, Youle MS, Johnson MA, Phillips A, Geretti AM. Prevalence and predictors of antiretroviral drug resistance in newly diagnosed HIV-1 infection. *J Antimicrob Chemother* 2007,**59**:517-524.

212. Fox J, Hill S, Kaye S, Dustan S, McClure M, Fidler S*, et al.* Prevalence of primary genotypic resistance in a UK centre: Comparison of primary HIV-1 and newly diagnosed treatment-naive individuals. *AIDS* 2007,**21**:237-239.

213. Dilernia DA, Lourtau L, Gomez AM, Ebenrstejin J, Toibaro JJ, Bautista CT*, et al.* Drug-resistance surveillance among newly HIV-1 diagnosed individuals in Buenos Aires, Argentina. *AIDS* 2007,**21**:1355-1360.

214. Barreto CC, Nishyia A, Araujo LV, Ferreira JE, Busch MP, Sabino EC. Trends in antiretroviral drug resistance and clade distributions among HIV-1--infected blood donors in Sao Paulo, Brazil. *J Acquir Immune Defic Syndr* 2006,**41**:338-341.

215. Varella RB, Ferreira SB, de Castro MB, Zalis MG, Tavares MD. Human immunodeficiency virus type 1 protease and reverse transcriptase mutation patterns among treatment-naive patients in different stages of infection in Rio de Janeiro, Brazil. *J Med Virol* 2007,**79**:1033-1039.

216. Perez L, Kouri V, Aleman Y, Abrahantes Y, Correa C, Aragones C*, et al.* Antiretroviral drug resistance in HIV-1 therapy-naive patients in Cuba. *Infect Genet Evol* 2013,**16**:144-150.

217. Murillo W, Lorenzana de Rivera I, Albert J, Guardado ME, Nieto AI, Paz-Bailey G. Prevalence of transmitted HIV-1 drug resistance among female sex workers and men who have sex with men in El Salvador, Central America. *J Med Virol* 2012,**84**:1514-1521.

218. Viani RM, Hsia K, Hubbard P, Ruiz-Calderon J, Lozada R, Alvelais J*, et al.* Prevalence of primary HIV-1 drug resistance in pregnant women and in newly diagnosed adults at Tijuana General Hospital, Baja California, Mexico. *Int J STD AIDS* 2007,**18**:235-238.

219. Lama JR, Sanchez J, Suarez L, Caballero P, Laguna A, Sanchez JL*, et al.* Linking HIV and antiretroviral drug resistance surveillance in Peru: a model for a third-generation HIV sentinel surveillance. *J Acquir Immune Defic Syndr* 2006,**42**:501-505.

220. Routy JP, Machouf N, Edwardes MD, Brenner BG, Thomas R, Trottier B*, et al.* Factors associated with a decrease in the prevalence of drug resistance in newly HIV-1 infected individuals in Montreal. *AIDS* 2004,**18**:2305-2312.

221. Liu L, May S, Richman DD, Hecht FM, Markowitz M, Daar ES*, et al.* Comparison of algorithms that interpret genotypic HIV-1 drug resistance to determine the prevalence of transmitted drug resistance. *AIDS* 2008,**22**:835-839.

222. Eshleman SH, Husnik M, Hudelson S, Donnell D, Huang Y, Huang W*, et al.* Antiretroviral drug resistance, HIV-1 tropism, and HIV-1 subtype among men who have sex with men with recent HIV-1 infection. *Aids* 2007,**21**:1165-1174.

223. Yanik EL, Napravnik S, Hurt CB, Dennis A, Quinlivan EB, Sebastian J*, et al.* Prevalence of Transmitted Antiretroviral Drug Resistance Differs Between Acutely and Chronically HIV-Infected Patients. *J Acquir Immune Defic Syndr* 2012,**61**:258-262.

224. Shet A, Berry L, Mohri H, Mehandru S, Chung C, Kim A*, et al.* Tracking the prevalence of transmitted antiretroviral drug-resistant HIV-1: a decade of experience. *J Acquir Immune Defic Syndr* 2006,**41**:439-446.

225. Castor D, Low A, Evering T, Karmon S, Davis B, Figueroa A*, et al.* Transmitted drug resistance and phylogenetic relationships among acute and early HIV-1-infected individuals in New York City. *J Acquir Immune Defic Syndr* 2012,**61**:1-8.

226. Hamers RL, Wallis CL, Kityo C, Siwale M, Mandaliya K, Conradie F*, et al.* HIV-1 drug resistance in antiretroviral-naive individuals in sub-Saharan Africa after rollout of antiretroviral therapy: a multicentre observational study. *Lancet Infect Dis* 2011,**11**:750-759.

227. Vergne L, Diagbouga S, Kouanfack C, Aghokeng A, Butel C, Laurent C*, et al.* HIV-1 drug-resistance mutations among newly diagnosed patients before scaling-up programmes in Burkina Faso and Cameroon. *Antivir Ther* 2006,**11**:575-579.

228. Muwonga J, Edidi S, Butel C, Vidal N, Monleau M, Okenge A*, et al.* Resistance to antiretroviral drugs in treated and drug-naive patients in the Democratic Republic of Congo. *J Acquir Immune Defic Syndr* 2011,**57 Suppl 1**:S27-33.

229. Kassu A, Fujino M, Matsuda M, Nishizawa M, Ota F, Sugiura W. Molecular epidemiology of HIV type 1 in treatment-naive patients in north Ethiopia. *AIDS Res Hum Retroviruses* 2007,**23**:564-568.

230. Hassan AS, Mwaringa SM, Obonyo CA, Nabwera HM, Sanders EJ, Rinke de Wit TF*, et al.* Low prevalence of transmitted HIV type 1 drug resistance among antiretroviral-naive adults in a rural HIV clinic in Kenya. *AIDS Res Hum Retroviruses* 2013,**29**:129-135.

231. Derache A, Maiga AI, Traore O, Akonde A, Cisse M, Jarrousse B*, et al.* Evolution of genetic diversity and drug resistance mutations in HIV-1 among untreated patients from Mali between 2005 and 2006. *J Antimicrob Chemother* 2008,**62**:456-463.

232. Maiga AI, Fofana DB, Maiga AC, Diallo F, Ait-Arkoub Z, Daou F*, et al.* Transmitted antiretroviral drug resistance in newly HIV-infected and untreated patients in Segou and Bamako, Mali. *AIDS Res Hum Retroviruses* 2013,**29**:182-186.

233. Annaz HE, Recordon-Pinson P, Baba N, Sedrati O, Mrani S, Fleury H. Presence of drug resistance mutations among drug-naive patients in Morocco. *AIDS Res Hum Retroviruses* 2011,**27**:917-920.

234. Diop-Ndiaye H, Toure-Kane C, Leye N, Ngom-Gueye NF, Montavon C, Peeters M*, et al.* Antiretroviral drug resistance mutations in antiretroviral-naive patients from Senegal. *AIDS Res Hum Retroviruses* 2010,**26**:1133-1138.

235. Nwobegahay J, Selabe G, Ndjeka NO, Manhaeve C, Bessong PO. Low prevalence of transmitted genetic drug resistance in a cohort of HIV infected naive patients entering antiretroviral treatment programs at two sites in northern South Africa. *J Med Virol* 2012,**84**:1839-1843.

236. Ndembi N, Lyagoba F, Nanteza B, Kushemererwa G, Serwanga J, Katongole-Mbidde E*, et al.* Transmitted antiretroviral drug resistance surveillance among newly HIV type 1-diagnosed women attending an antenatal clinic in Entebbe, Uganda. *AIDS Res Hum Retroviruses* 2008,**24**:889-895.

237. Arora SK, Gupta S, Toor JS, Singla A. Drug resistance-associated genotypic alterations in the pol gene of HIV type 1 isolates in ART-naive individuals in North India. *AIDS Res Hum Retroviruses* 2008,**24**:125-130.

238. Chaturbhuj DN, Hingankar NK, Srikantiah P, Garg R, Kabra S, Deshmukh PS*, et al.* Transmitted HIV drug resistance among HIV-infected voluntary counseling and testing centers (VCTC) clients in Mumbai, India. *AIDS Res Hum Retroviruses* 2010,**26**:927-932.

239. Thorat SR, Chaturbhuj DN, Hingankar NK, Chandrasekhar V, Koppada R, Datkar SR*, et al.* Surveillance of transmitted HIV type 1 drug resistance among HIV type 1-positive women attending an antenatal clinic in Kakinada, India. *AIDS Res Hum Retroviruses* 2011,**27**:1291-1297.

240. Jahanbakhsh F, Hattori J, Matsuda M, Ibe S, Monavari SH, Memarnejadian A*, et al.* Prevalence of transmitted HIV drug resistance in Iran between 2010 and 2011. *PLoS ONE* 2013,**8**:e61864.

241. Grossman Z, Lorber M, Maayan S, Bar-Yacov N, Levy I, Averbuch D*, et al.* Drug-resistant HIV infection among drug-naive patients in Israel. *Clin Infect Dis* 2005,**40**:294-302.

242. Ibe S, Hattori J, Fujisaki S, Shigemi U, Fujisaki S, Shimizu K*, et al.* Trend of drug-resistant HIV type 1 emergence among therapy-naive patients in Nagoya, Japan:

an 8-year surveillance from 1999 to 2006. *AIDS Res Hum Retroviruses* 2008,**24**:7-14.

243. Hattori J, Shiino T, Gatanaga H, Yoshida S, Watanabe D, Minami R*, et al.* Trends in transmitted drug-resistant HIV-1 and demographic characteristics of newly diagnosed patients: nationwide surveillance from 2003 to 2008 in Japan. *Antiviral Res* 2010,**88**:72-79.

244. Bang JI, Song KH, Kim SH, Cho JH, Park WB, Park SW*, et al.* Prevalence of primary antiretroviral resistance: trends in Korea. *AIDS Res Hum Retroviruses* 2008,**24**:83-85.

245. Tee KK, Kamarulzaman A, Ng KP. Short communication: low prevalence of genotypic drug resistance mutations among antiretroviral-naive HIV type 1 patients in Malaysia. *AIDS Res Hum Retroviruses* 2006,**22**:121-124.

246. Chang SY, Chen MY, Lee CN, Sun HY, Ko W, Chang SF*, et al.* Trends of antiretroviral drug resistance in treatment-naive patients with human immunodeficiency virus type 1 infection in Taiwan. *J Antimicrob Chemother* 2008,**61**:689-693.

247. Apisarnthanarak A, Mundy LM. Antiretroviral drug resistance among antiretroviral-naive individuals with HIV infection of unknown duration in Thailand. *Clin Infect Dis* 2008,**46**:1630-1631.

248. Sungkanuparph S, Sukasem C, Kiertiburanakul S, Pasomsub E, Chantratita W. Emergence of HIV-1 drug resistance mutations among antiretroviral-naive HIV-1-infected patients after rapid scaling up of antiretroviral therapy in Thailand. *J Int AIDS Soc* 2012,**15**:12.

249. Manosuthi W, Thongyen S, Nilkamhang S, Manosuthi S, Sungkanuparph S. HIV-1 drug resistance-associated mutations among antiretroviral-naive Thai patients with chronic HIV-1 infection. *J Med Virol* 2013,**85**:194-199.

250. Duc NB, Hien BT, Wagar N, Tram TH, Giang le T, Yang C*, et al.* Surveillance of transmitted HIV drug resistance using matched plasma and dried blood spot specimens from voluntary counseling and testing sites in Ho Chi Minh City, Vietnam, 2007-2008. *Clin Infect Dis* 2012,**54 Suppl 4**:S343-347.

251. Dean J, Ta Thi TH, Dunford L, Carr MJ, Nguyen LT, Coughlan S*, et al.* Prevalence of HIV type 1 antiretroviral drug resistance mutations in Vietnam: a multicenter study. *AIDS Res Hum Retroviruses* 2011,**27**:797-801.

252. Bannister WP, Cozzi-Lepri A, Clotet B, Mocroft A, Kjaer J, Reiss P*, et al.* Transmitted drug resistant HIV-1 and association with virologic and CD4 cell count response to combination antiretroviral therapy in the EuroSIDA Study. *J Acquir Immune Defic Syndr* 2008,**48**:324-333.

253. Vercauteren J, Derdelinckx I, Sasse A, Bogaert M, Ceunen H, De Roo A*, et al.* Prevalence and epidemiology of HIV type 1 drug resistance among newly diagnosed therapy-naive patients in Belgium from 2003 to 2006. *AIDS Res Hum Retroviruses* 2008,**24**:355-362.

254. Grgic I, Lepej SZ, Lunar MM, Poljak M, Vince A, Vrakela IB*, et al.* The prevalence of transmitted drug resistance in newly diagnosed HIV-infected individuals in Croatia: the role of transmission clusters of men who have sex with men carrying the T215S surveillance drug resistance mutation. *AIDS Res Hum Retroviruses* 2013,**29**:329-336.

255. Audelin AM, Gerstoft J, Obel N, Mathiesen L, Laursen A, Pedersen C*, et al.* Molecular phylogenetics of transmitted drug resistance in newly diagnosed HIV Type 1 individuals in Denmark: a nation-wide study. *AIDS Res Hum Retroviruses* 2011,**27**:1283-1290.

256. Vazquez de Parga E, Rakhmanova A, Perez-Alvarez L, Vinogradova A, Delgado E, Thomson MM*, et al.* Analysis of drug resistance-associated mutations in treatment-naive individuals infected with different genetic forms of HIV-1 circulating in countries of the former Soviet Union. *J Med Virol* 2005,**77**:337-344.

257. Sagir A, Oette M, Kaiser R, Daumer M, Fatkenheuer G, Rockstroh JK*, et al.* Trends of prevalence of primary HIV drug resistance in Germany. *J Antimicrob Chemother* 2007,**60**:843-848.

258. Paraskevis D, Magiorkinis E, Katsoulidou A, Hatzitheodorou E, Antoniadou A, Papadopoulos A*, et al.* Prevalence of resistance-associated mutations in newly diagnosed HIV-1 patients in Greece. *Virus Res* 2005,**112**:115-122.

259. Skoura L, Metallidis S, Pilalas D, Kourelis A, Margariti A, Papadimitriou E*, et al.* High rates of transmitted drug resistance among newly-diagnosed antiretroviral naive HIV patients in Northern Greece, data from 2009-2011. *Clin Microbiol Infect* 2013,**19**:E169-172.

260. Bracciale L, Colafigli M, Zazzi M, Corsi P, Meraviglia P, Micheli V*, et al.* Prevalence of transmitted HIV-1 drug resistance in HIV-1-infected patients in Italy: evolution over 12 years and predictors. *J Antimicrob Chemother* 2009,**64**:607-615.

261. Lapadula G, Izzo I, Gargiulo F, Paraninfo G, Castelnuovo F, Quiros-Roldan E*, et al.* Updated prevalence of genotypic resistance among HIV-1 positive patients naive to antiretroviral therapy: a single center analysis. *J Med Virol* 2008,**80**:747-753.

262. Balode D, Westman M, Kolupajeva T, Rozentale B, Albert J. Low prevalence of transmitted drug resistance among newly diagnosed HIV-1 patients in Latvia. *J Med Virol* 2010,**82**:2013-2018.

263. Babic DZ, Zelnikar M, Seme K, Vandamme AM, Snoeck J, Tomazic J*, et al.* Prevalence of antiretroviral drug resistance mutations and HIV-1 non-B subtypes in newly diagnosed drug-naive patients in Slovenia, 2000-2004. *Virus Res* 2006,**118**:156-163.

264. Garcia F, Perez-Cachafeiro S, Guillot V, Alvarez M, Perez-Romero P, Perez-Elias MJ*, et al.* Transmission of HIV drug resistance and non-B subtype distribution in the Spanish cohort of antiretroviral treatment naive HIV-infected individuals (CoRIS). *Antiviral Res* 2011,**91**:150-153.

265. Monge S, Guillot V, Alvarez M, Pena A, Viciana P, Garcia-Bujalance S*, et al.* Analysis of transmitted drug resistance in Spain in the years 2007-2010 documents a decline in mutations to the non-nucleoside drug class. *Clin Microbiol Infect* 2012,**18**:E485-490.

266. Karlsson A, Bjorkman P, Bratt G, Ekvall H, Gisslen M, Sonnerborg A*, et al.* Low prevalence of transmitted drug resistance in patients newly diagnosed with HIV-1 infection in Sweden 2003-2010. *PLoS ONE* 2012,**7**:e33484.

267. Green H, Tilston P, Fearnhill E, Pillay D, Dunn DT. The Impact of Different Definitions on the Estimated Rate of Transmitted HIV Drug Resistance in the United Kingdom. *J Acquir Immune Defic Syndr* 2008.

268. Payne BA, Nsutebu EF, Hunter ER, Olarinde O, Collini P, Dunbar JA*, et al.* Low prevalence of transmitted antiretroviral drug resistance in a large UK HIV-1 cohort. *J Antimicrob Chemother* 2008,**62**:464-468.

269. Dolling D, Sabin C, Delpech V, Smit E, Pozniak A, Asboe D*, et al.* Time trends in drug resistant HIV-1 infections in the United Kingdom up to 2009: multicentre observational study. *BMJ* 2012,**345**:e5253.

270. Rodrigues R, Scherer LC, Oliveira CM, Franco HM, Sperhacke RD, Ferreira JL*, et al.* Low prevalence of primary antiretroviral resistance mutations and

predominance of HIV-1 clade C at polymerase gene in newly diagnosed individuals from south Brazil. *Virus Res* 2006,**116**:201-207.

271. Arruda E, Simoes L, Sucupira C, Medeiros M, Diaz RS, Lima A. Short communication: intermediate prevalence of HIV type 1 primary antiretroviral resistance in Ceara State, Northeast Brazil. *AIDS Res Hum Retroviruses* 2011,**27**:153-156.

272. Bermudez-Aza EH, Kerr LR, Kendall C, Pinho AA, de Mello MB, Mota RS*, et al.* Antiretroviral drug resistance in a respondent-driven sample of HIV-infected men who have sex with men in Brazil. *J Acquir Immune Defic Syndr* 2011,**57 Suppl 3**:S186-192.

273. Graf T, Passaes CP, Ferreira LG, Grisard EC, Morgado MG, Bello G*, et al.* HIV-1 genetic diversity and drug resistance among treatment naive patients from Southern Brazil: an association of HIV-1 subtypes with exposure categories. *J Clin Virol* 2011,**51**:186-191.

274. Rios M, Delgado E, Perez-Alvarez L, Fernandez J, Galvez P, de Parga EV*, et al.* Antiretroviral drug resistance and phylogenetic diversity of HIV-1 in Chile. *J Med Virol* 2007,**79**:647-656.

275. Myers JE, Taylor BS, Rojas Fermin RA, Reyes EV, Vaughan C, Jose L*, et al.* Transmitted drug resistance among antiretroviral-naive patients with established HIV type 1 infection in Santo Domingo, Dominican Republic and review of the Latin American and Caribbean literature. *AIDS Res Hum Retroviruses* 2012,**28**:667-674.

276. Lloyd B, O'Connell RJ, Michael NL, Aviles R, Palou E, Hernandez R*, et al.* Prevalence of resistance mutations in HIV-1-Infected Hondurans at the beginning of the National Antiretroviral Therapy Program. *AIDS Res Hum Retroviruses* 2008,**24**:529-535.

277. Murillo W, Paz-Bailey G, Morales S, Monterroso E, Paredes M, Dobbs T*, et al.* Transmitted drug resistance and type of infection in newly diagnosed HIV-1 individuals in Honduras. *J Clin Virol* 2010,**49**:239-244.

278. Escoto-Delgadillo M, Vazquez-Valls E, Ramirez-Rodriguez M, Corona-Nakamura A, Amaya-Tapia G, Quintero-Perez N*, et al.* Drug-resistance mutations in antiretroviral-naive patients with established HIV-1 infection in Mexico. *HIV Med* 2005,**6**:403-409.

279. Ross L, Lim ML, Liao Q, Wine B, Rodriguez AE, Weinberg W*, et al.* Prevalence of antiretroviral drug resistance and resistance-associated mutations in antiretroviral therapy-naive HIV-infected individuals from 40 United States cities. *HIV Clin Trials* 2007,**8**:1-8.

280. Poon AF, Aldous JL, Mathews WC, Kitahata M, Kahn JS, Saag MS*, et al.* Transmitted drug resistance in the CFAR network of integrated clinical systems cohort: prevalence and effects on pre-therapy CD4 and viral load. *PLoS ONE* 2011,**6**:e21189.

281. Hightow-Weidman LB, Hurt CB, Phillips G, 2nd, Jones K, Magnus M, Giordano TP*, et al.* Transmitted HIV-1 drug resistance among young men of color who have sex with men: a multicenter cohort analysis. *J Adolesc Health* 2011,**48**:94-99.

282. Wheeler WH, Ziebell RA, Zabina H, Pieniazek D, Prejean J, Bodnar UR*, et al.* Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S.-2006. *AIDS* 2010,**24**:1203-1212.

283. Huang HY, Daar ES, Sax PE, Young B, Cook P, Benson P*, et al.* The prevalence of transmitted antiretroviral drug resistance in treatment-naive patients and factors influencing first-line treatment regimen selection. *HIV Med* 2008,**9**:285-293.

284. Klibanov OM, Dolder CR, Assefa F, Ekwonu TJ. Baseline antiretroviral resistance and clinical outcomes in an urban HIV clinic. *AIDS Patient Care STDS* 2013,**27**:205-207.

285. Huaman MA, Aguilar J, Baxa D, Golembieski A, Brar I, Markowitz N. Late presentation and transmitted drug resistance mutations in new HIV-1 diagnoses in Detroit. *Int J Infect Dis* 2011,**15**:e764-768.

286. MacVeigh MS, Kosmetatos MK, McDonald JE, Reeder JL, Parrish DA, Young TP. Prevalence of drug-resistant HIV type 1 at the time of initiation of antiretroviral therapy in Portland, Oregon. *AIDS Res Hum Retroviruses* 2013,**29**:337-342.

287. Grubb JR, Singhatiraj E, Mondy K, Powderly WG, Overton ET. Patterns of primary antiretroviral drug resistance in antiretroviral-naive HIV-1-infected individuals in a midwest university clinic. *AIDS* 2006,**20**:2115-2116.

288. Taniguchi T, Nurutdinova D, Grubb JR, Onen NF, Shacham E, Donovan M*, et al.* Transmitted drug-resistant HIV type 1 remains prevalent and impacts virologic

outcomes despite genotype-guided antiretroviral therapy. *AIDS Res Hum Retroviruses* 2012,**28**:259-264.

289. Boyd AC, Herzberg EM, Marshall MM, Lamparello NA, De Leon MA, Porter A*, et al.* Antiretroviral Drug Resistance Among Treatment-Naive HIV-1-Infected Persons in Washington, D.C. *AIDS Patient Care STDS* 2008.

290. Readhead AC, Gordon DE, Wang Z, Anderson BJ, Brousseau KS, Kouznetsova MA*, et al.* Transmitted antiretroviral drug resistance in New York State, 2006-2008: results from a new surveillance system. *PLoS ONE* 2012,**7**:e40533.

291. Wensing AM, van de Vijver DA, Angarano G, Asjo B, Balotta C, Boeri E*, et al.* Prevalence of drug-resistant HIV-1 variants in untreated individuals in Europe: implications for clinical management. *J Infect Dis* 2005,**192**:958-966.

292. Frentz D, Boucher CA, van de Vijver DA. Temporal changes in the epidemiology of transmission of drug-resistant HIV-1 across the world. *AIDS Rev* 2012,**14**:17-27.

293. Bertagnolio S, Derdelinckx I, Parker M, Fitzgibbon J, Fleury H, Peeters M*, et al.* World Health Organization/HIVResNet Drug Resistance Laboratory Strategy. *Antivir Ther* 2008,**13 Suppl 2**:49-57.

294. Gupta RK, Jordan MR, Sultan BJ, Hill A, Davis DH, Gregson J*, et al.* Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. *Lancet* 2012,**380**:1250-1258.

295. Corvasce S, Violin M, Romano L, Razzolini F, Vicenti I, Galli A*, et al.* Evidence of differential selection of HIV-1 variants carrying drug-resistant mutations in seroconverters. *Antivir Ther* 2006,**11**:329-334.

296. Yerly S, Jost S, Telenti A, Flepp M, Kaiser L, Chave JP*, et al.* Infrequent transmission of HIV-1 drug-resistant variants. *Antivir Ther* 2004,**9**:375-384.

297. Turner D, Brenner B, Routy JP, Moisi D, Rosberger Z, Roger M*, et al.* Diminished representation of HIV-1 variants containing select drug resistance-conferring mutations in primary HIV-1 infection. *J Acquir Immune Defic Syndr* 2004,**37**:1627-1631.

298. Bezemer D, de Ronde A, Prins M, Porter K, Gifford R, Pillay D*, et al.* Evolution of transmitted HIV-1 with drug-resistance mutations in the absence of therapy: effects on CD4+ T-cell count and HIV-1 RNA load. *Antivir Ther* 2006,**11**:173-178.

299. Pao D, Andrady U, Clarke J, Dean G, Drake S, Fisher M*, et al.* Long-term persistence of primary genotypic resistance after HIV-1 seroconversion. *J Acquir Immune Defic Syndr* 2004,**37**:1570-1573.

300. Bleiber G, Munoz M, Ciuffi A, Meylan P, Telenti A. Individual contributions of mutant protease and reverse transcriptase to viral infectivity, replication, and protein maturation of antiretroviral drug-resistant human immunodeficiency virus type 1. *J Virol* 2001,**75**:3291-3300.

301. Cong ME, Heneine W, Garcia-Lerma JG. The fitness cost of mutations associated with human immunodeficiency virus type 1 drug resistance is modulated by mutational interactions. *J Virol* 2007,**81**:3037-3041.

302. Bezemer D, Jurriaans S, Prins M, van der Hoek L, Prins JM, de Wolf F*, et al.* Declining trend in transmission of drug-resistant HIV-1 in Amsterdam. *AIDS* 2004,**18**:1571-1577.

303. Poggensee G, Kucherer C, Werning J, Somogyi S, Bieniek B, Dupke S*, et al.* Impact of transmission of drug-resistant HIV on the course of infection and the treatment success. Data from the German HIV-1 Seroconverter Study. *HIV Med* 2007,**8**:511-519.

304. Pillay D, Bhaskaran K, Jurriaans S, Prins M, Masquelier B, Dabis F*, et al.* The impact of transmitted drug resistance on the natural history of HIV infection and response to first-line therapy. *AIDS* 2006,**20**:21-28.

305. Peuchant O, Thiebaut R, Capdepont S, Lavignolle-Aurillac V, Neau D, Morlat P*, et al.* Transmission of HIV-1 minority-resistant variants and response to first-line antiretroviral therapy. *AIDS* 2008,**22**:1417-1423.

306. Tamalet C, Pasquier C, Yahi N, Colson P, Poizot-Martin I, Lepeu G*, et al.* Prevalence of drug resistant mutants and virological response to combination therapy in patients with primary HIV-1 infection. *J Med Virol* 2000,**61**:181-186.

307. 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.

308. Shafer RW, Warford A, Winters MA, Gonzales MJ. Reproducibility of human immunodeficiency virus type 1 (HIV-1) protease and reverse transcriptase sequencing of plasma samples from heavily treated HIV-1-infected individuals. *J Virol Methods* 2000,**86**:143-153.

309. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977,**74**:5463-5467.

310. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR*, et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* 1986,**321**:674-679.

311. Winters B, Van Craenenbroeck E, Van der Borght K, Lecocq P, Villacian J, Bacheler L. Clinical cut-offs for HIV-1 phenotypic resistance estimates: update based on recent pivotal clinical trial data and a revised approach to viral mixtures. *J Virol Methods* 2009,**162**:101-108.

312. Meynard JL, Vray M, Morand-Joubert L, Race E, Descamps D, Peytavin G*, et al.* Phenotypic or genotypic resistance testing for choosing antiretroviral therapy after treatment failure: a randomized trial. *AIDS* 2002,**16**:727-736.

313. Johnson VA, Calvez V, Gunthard HF, Paredes R, Pillay D, Shafer R*, et al.* 2011 update of the drug resistance mutations in HIV-1. *Top Antivir Med* 2011,**19**:156-164.

314. Kutyavin IV, Afonina IA, Mills A, Gorn VV, Lukhtanov EA, Belousov ES*, et al.* 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res* 2000,**28**:655-661.

315. Metzner KJ, Bonhoeffer S, Fischer M, Karanicolas 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. *J Infect Dis* 2003,**188**:1433-1443.

316. 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. *J Clin Microbiol* 2006,**44**:2612-2614.

317. Edelstein RE, Nickerson DA, Tobe VO, Manns-Arcuino LA, Frenkel LM. Oligonucleotide ligation assay for detecting mutations in the human immunodeficiency virus type 1 pol gene that are associated with resistance to zidovudine, didanosine, and lamivudine. *J Clin Microbiol* 1998,**36**:569-572.

318. Kapoor A, Jones M, Shafer RW, Rhee SY, Kazanjian P, Delwart EL. Sequencing-based detection of low-frequency human immunodeficiency virus type 1 drug-resistant mutants by an RNA/DNA heteroduplex generator-tracking assay. *J Virol* 2004,**78**:7112-7123.

319. Resch W, Parkin N, Stuelke EL, Watkins T, Swanstrom R. A multiple-site-specific heteroduplex tracking assay as a tool for the study of viral population dynamics. *Proc Natl Acad Sci U S A* 2001,**98**:176-181.

320. Schnell G, Ince WL, Swanstrom R. Identification and recovery of minor HIV-1 variants using the heteroduplex tracking assay and biotinylated probes. *Nat Methods (in press)* 2008.

321. Holterman L, Dubbes R, Mullins J, Haaijman J, Heeney J. A strategy for cloning infectious molecular clones of retroviruses from serum or plasma. *J Virol Methods* 2000,**84**:37-48.

322. Paolucci S, Baldanti F, Campanini G, Zavattoni M, Cattaneo E, Dossena L*, et al.* Analysis of HIV drug-resistant quasispecies in plasma, peripheral blood mononuclear cells and viral isolates from treatment-naive and HAART patients. *J Med Virol* 2001,**65**:207-217.

323. Malet I, Belnard M, Agut H, Cahour A. From RNA to quasispecies: a DNA polymerase with proofreading activity is highly recommended for accurate assessment of viral diversity. *J Virol Methods* 2003,**109**:161-170.

324. Condra JH, Schleif WA, Blahy OM, Gabryelski LJ, Graham DJ, Quintero JC*, et al.* In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* 1995,**374**:569-571.

325. O'Meara D, Wilbe K, Leitner T, Hejdeman B, Albert J, Lundeberg J. Monitoring resistance to human immunodeficiency virus type 1 protease inhibitors by pyrosequencing. *J Clin Microbiol* 2001,**39**:464-473.

326. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG*, et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008,**456**:53-59.

327. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA*, et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005,**437**:376-380.

328. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A* 2003,**100**:8817-8822.

329. Leamon JH, Lee WL, Tartaro KR, Lanza JR, Sarkis GJ, deWinter AD*, et al.* A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. *Electrophoresis* 2003,**24**:3769-3777.

330. Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I*, et al.* Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc Natl Acad Sci U S A* 2008,**105**:13081-13086.

331. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet* 2010,**11**:31-46.

332. Holodniy M, Mole L, Yen-Lieberman B, Margolis D, Starkey C, Carroll R*, et al.* Comparative stabilities of quantitative human immunodeficiency virus RNA in plasma from samples collected in VACUTAINER CPT, VACUTAINER PPT, and standard VACUTAINER tubes. *J Clin Microbiol* 1995,**33**:1562-1566.

333. Cepko C. Large-scale preparation and concentration of retrovirus stocks. *Curr Protoc Mol Biol* 2001,**Chapter 9**:Unit9 12.

334. Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science* 1988,**242**:1171-1173.

335. Fang G, Zhu G, Burger H, Keithly JS, Weiser B. Minimizing DNA recombination during long RT-PCR. *J Virol Methods* 1998,**76**:139-148.

336. Metzner KJ, Rauch P, Walter H, Boesecke C, Zollner B, Jessen H*, et al.* Detection of minor populations of drug-resistant HIV-1 in acute seroconverters. *AIDS* 2005,**19**:1819-1825.

337. Cline J, Braman JC, Hogrefe HH. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res* 1996,**24**:3546-3551.

338. Yang YL, Wang G, Dorman K, Kaplan AH. Long polymerase chain reaction amplification of heterogeneous HIV type 1 templates produces recombination at a relatively high frequency. *AIDS Res Hum Retroviruses* 1996,**12**:303-306.

339. Napravnik S, Eron JJ, Jr., McKaig RG, Heine AD, Menezes P, Quinlivan E. Factors associated with fewer visits for HIV primary care at a tertiary care center in the Southeastern U.S. *AIDS Care* 2006,**18 Suppl 1**:S45-50.

340. Adimora AA, Schoenbach VJ, Doherty IA. HIV and African Americans in the southern United States: sexual networks and social context. *Sex Transm Dis* 2006,**33**:S39-45.

341. Karon JM, Fleming PL, Steketee RW, De Cock KM. HIV in the United States at the turn of the century: an epidemic in transition. *Am J Public Health* 2001,**91**:1060-1068.

342. Dennis AM, Napravnik S, Sena AC, Eron JJ. Late entry to HIV care among Latinos compared with non-Latinos in a southeastern US cohort. *Clin Infect Dis* 2011,**53**:480-487.

343. Yanik EL, Tamburro K, Eron JJ, Damania B, Napravnik S, Dittmer DP. Recent cancer incidence trends in an observational clinical cohort of HIV-infected patients in the US, 2000 to 2011. *Infect Agent Cancer* 2013,**8**:18.

344. University of North Carolina. UNC Center for AIDS Research web site. In; 2007.

345. Jose M, Gajardo R, Jorquera JI. Stability of HCV, HIV-1 and HBV nucleic acids in plasma samples under long-term storage. *Biologicals* 2005,**33**:9-16.

346. Simen BB, Simons JF, Hullsiek KH, Novak RM, Macarthur RD, Baxter JD*, et al.* Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. *J Infect Dis* 2009,**199**:693-701.

347. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG*, et al.* Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 2003,**31**:3497-3500.

348. 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. *Mol Biol Evol* 2011,**28**:2731-2739.

349. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980,**16**:111-120.

350. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* 2007,**8**:R143.

351. Williams RL. A note on robust variance estimation for cluster-correlated data. *Biometrics* 2000,**56**:645-646.

352. Cox DR. Regression Models and Life-Tables. *J R Stat Soc [Ser B]* 1972,**34**:187-220.

353. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 2nd ed. Hoboken, N.J.: J. Wiley; 2002.

354. Efron B. The Efficiency of Cox's Likelihood Function for Censored Data. *J Am Stat Assoc* 1977,**72**:557-565.

355. Breslow N. Covariance analysis of censored survival data. *Biometrics* 1974,**30**:89-99.

356. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet* 2002,**359**:1686-1689.

357. Allison PD, SAS Institute. "Heterogeneity, repeated events, and other topics". In: *Survival Analysis Using the SAS system: A Practical Guide*. Cary, NC: SAS Institute; 1995. pp. viii, 292 p.249-252.

358. Buchacz K, Baker RK, Palella FJ, Jr., Chmiel JS, Lichtenstein KA, Novak RM*, et al.* AIDS-defining opportunistic illnesses in US patients, 1994-2007: a cohort study. *AIDS* 2010,**24**:1549-1559.

359. Moore RD, Keruly JC, Bartlett JG. Improvement in the Health of HIV-Infected Persons in Care: Reducing Disparities. *Clin Infect Dis* 2012,**55**:1242-1251.

360. Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, Matte C*, et al.* High rates of forward transmission events after acute/early HIV-1 infection. *J Infect Dis* 2007,**195**:951-959.

361. Yerly S, Junier T, Gayet-Ageron A, Amari EB, von Wyl V, Gunthard HF*, et al.* The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. *AIDS* 2009,**23**:1415-1423.

362. Bansi L, Geretti AM, Dunn D, Hill T, Green H, Fearnhill E*, et al.* Impact of transmitted drug-resistance on treatment selection and outcome of first-line Highly Active Antiretroviral Therapy (HAART). *J Acquir Immune Defic Syndr* 2010,**53**:633-639.

363. Wittkop L, Gunthard HF, de Wolf F, Dunn D, Cozzi-Lepri A, de Luca A*, et al.* Effect of transmitted drug resistance on virological and immunological response to initial combination antiretroviral therapy for HIV (EuroCoord-CHAIN joint project): a European multicohort study. *Lancet Infect Dis* 2011,**11**:363-371.

364. Goodman DD, Zhou Y, Margot NA, McColl DJ, Zhong L, Borroto-Esoda K*, et al.* Low level of the K103N HIV-1 above a threshold is associated with virological failure in treatment-naive individuals undergoing efavirenz-containing therapy. *AIDS* 2011,**25**:325-333.

365. Jakobsen MR, Tolstrup M, Sogaard OS, Jorgensen LB, Gorry PR, Laursen A*, et al.* Transmission of HIV-1 drug-resistant variants: prevalence and effect on treatment outcome. *Clin Infect Dis* 2010,**50**:566-573.

366. Johnson JA, Li JF, Wei X, Craig C, Stone C, Horton JH*, et al.* Baseline detection of low-frequency drug resistance-associated mutations is strongly associated with virological failure in previously antiretroviral-naive HIV-1-infected persons. *Antivir Ther* 2006,**11**:S79-S79.

367. Lataillade M, Chiarella J, Yang R, Schnittman S, Wirtz V, Uy J*, et al.* Prevalence and clinical significance of HIV drug resistance mutations by ultra-deep sequencing in antiretroviral-naive subjects in the CASTLE study. *PLoS ONE* 2010,**5**:e10952.

368. Metzner KJ, Rauch P, Braun P, Knechten H, Ehret R, Korn K*, et al.* Prevalence of key resistance mutations K65R, K103N, and M184V as minority HIV-1 variants in chronically HIV-1 infected, treatment-naive patients. *J Clin Virol* 2011,**50**:156-161.

369. Varghese V, Wang E, Babrzadeh F, Bachmann MH, Shahriar R, Liu T*, et al.* Nucleic acid template and the risk of a PCR-Induced HIV-1 drug resistance mutation. *PLoS ONE* 2010,**5**:e10992.

370. Kuritzkes DR, Lalama CM, Ribaudo HJ, Marcial M, Meyer WA, 3rd, Shikuma C*, et al.* Preexisting resistance to nonnucleoside reverse-transcriptase inhibitors predicts virologic failure of an efavirenz-based regimen in treatment-naive HIV-1-infected subjects. *J Infect Dis* 2008,**197**:867-870.

371. Krueger F, Andrews SR, Osborne CS. Large scale loss of data in low-diversity illumina sequencing libraries can be recovered by deferred cluster calling. *PLoS ONE* 2011,**6**:e16607.

372. Wang J, Bambara RA, Demeter LM, Dykes C. Reduced fitness in cell culture of HIV-1 with nonnucleoside reverse transcriptase inhibitor-resistant mutations correlates with relative levels of reverse transcriptase content and RNase H activity in virions. *J Virol* 2010,**84**:9377-9389.

373. Wang J, Li D, Bambara RA, Yang H, Dykes C. L74V increases the reverse transcriptase content of HIV-1 virions with non-nucleoside reverse transcriptase drug-resistant mutations L100I+K103N and K101E+G190S, which results in increased fitness. *J Gen Virol* 2013,**94**:1597-1607.

374. Stekler JD, Ellis GM, Carlsson J, Eilers B, Holte S, Maenza J*, et al.* Prevalence and impact of minority variant drug resistance mutations in primary HIV-1 infection. *PLoS ONE* 2011,**6**:e28952.