

Department of Microbiology, Tumor and Cell Biology
Karolinska Institutet, Stockholm, Sweden

ULTRA-DEEP CHARACTERIZATION OF VIRAL QUASISPECIES IN HIV INFECTION

Charlotte Hedskog



**Karolinska
Institutet**

Stockholm 2012

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet. Printed by Larserics Digital print AB

© Charlotte Hedskog, 2012
ISBN 978-91-7457-772-3

To my family

ABSTRACT

HIV-1 has the ability to rapidly diversify and adapt to changes in its environment, such as evading the host immune response, altering cell tropism, and developing resistance to antiretroviral drugs. Minority HIV-1 variants have been shown to be of clinical significance, especially those with non-nucleoside reverse transcriptase inhibitor (NNRTI) drug resistance mutations or determinants of CXCR4 phenotype (X4-virus). In this thesis a next generation sequencing technology, ultra-deep pyrosequencing (UDPS), has been used to dissect HIV-1 quasispecies in infected patients to study the evolution of drug resistance and cell tropism. The depth of UDPS depends on the number of viral templates that can be successfully extracted and amplified from a plasma sample, the error rate of PCR and UDPS, and the efficiency of cleaning the UDPS data from such errors. For this reason, we developed an experimental design that allows high recovery of HIV-1 templates and an efficient data cleaning strategy. Our data cleaning strategy reduced the UDPS error rate approximately 10-fold. We carefully evaluated the performance of our UDPS protocol and found that the repeatability of detection of major as well as minor variants in patient plasma samples was good. This indicated that the experimental noise introduced during RNA extraction, cDNA synthesis, PCR and UDPS was low. However, for rare variants *in vitro* PCR recombination and effects of sequence direction need to be considered. Finally, the design of primers for PCR amplification is of special importance during UDPS, since we observed that primer-related selective amplification can skew the frequency estimates of genetic variants.

We investigated the levels of pre-existing drug resistance mutations in plasma samples from five treatment-naïve patients. In four of five patients we found low levels of pre-existing drug resistance mutations at two positions (M184I, T215A/I), whereas other mutations (M184V, Y181C, Y188C and T215Y/F) were not detected. During treatment failure and treatment interruption, we found almost complete replacement of wild-type and drug-resistant variants, respectively. This implies that the proportion of minority variants with drug resistance in patients with previous treatment failure or transmitted drug resistance can be too low to be detectable even with highly sensitive UDPS. In another study, the HIV-1 populations from three patients with HIV-1 populations that switched coreceptor use were investigated longitudinally. UDPS analysis showed that the X4-virus that emerged after coreceptor switch was not detected during primary HIV-1 infection (PHI) and that the X4 population most probably evolved from the CCR5-using population during the course of infection rather than was transmitted as minor variants. Moreover, one to three major variants were found during PHI, supporting that infection usually is established with one or just a few viral particles.

The frequency and type of errors that occurred during UDPS were investigated. The errors that remained after data cleaning were significantly more often transitions than transversions, which indicates that a substantial proportion of errors were introduced during PCR rather than UDPS itself. This affects the limits of detection of minority mutations since UDPS analyses of HIV-1 are preceded by a PCR step. To further reduce the UDPS error rate we developed a new, improved

methodology, based on re-sequencing of molecularly tagged template molecules. Preliminary results showed that this method has the potential to increase the sensitivity of UDPS analyses 1000-fold and thus is close to error-free.

Taken together, this thesis adds knowledge on the use of UDPS to gain new insights in HIV evolution and resistance and is relevant for the possible future clinical use of this technology.

LIST OF PUBLICATIONS

This thesis is based on the following papers, referred to in the text by their Roman numerals:

- I. **Hedskog C**, Mild M, Jernberg J, Sherwood E, Bratt G, Leitner T, Lundeberg J, Andersson B, Albert J. Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. PLoS One. 2010 Jul 7;5(7):e11345.
- II. Mild M*, **Hedskog C***, Jernberg J, Albert J. Performance of ultra-deep pyrosequencing in analysis of HIV-1 *pol* gene variation. PLoS One. 2011;6(7):e22741. Epub 2011 Jul 25. (Shared first authorship).
- III. **Hedskog C**, Brodin J, Heddini A, Bratt G, Albert J, Mild M. Longitudinal ultra-deep characterization of HIV-1 R5 and X4 subpopulations in patients followed from primary infection to coreceptor switch.
- IV. Brodin J, Mild M, **Hedskog C**, Sherwood E, Leitner T, Andersson B, Albert J. Sources and characteristics of errors in ultra-deep pyrosequencing and development of a data cleaning strategy.

LIST OF ABBREVIATIONS

AIDS	Acquired Immunodeficiency Syndrome
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
cART	Combination antiretroviral therapy
CCR5	C-C chemokine receptor type 5
CD4	Cluster of differentiation 4
cDNA	Complementary DNA
CTL	Cytotoxic T-lymphocyte
CXCR4	C-X-C motif receptor 4
DNA	Deoxyribonucleic acid
Em-PCR	Emulsion based PCR
Env	Envelope
Gag	Group specific antigen
HAART	Highly active antiretroviral therapy
HIV-1	Human immunodeficiency virus type 1
HIV-2	Human immunodeficiency virus type 2
HLA	Human leukocyte antigen
IN	Integrase
LTR	Long terminal repeat
MHC	Major histocompatibility complex
mRNA	Messenger RNA
MSM	Men who have sex with men
Nef	Negative factor
NGS	Next generation sequencing
NJ	Neighbor joining
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside reverse transcriptase inhibitor
PBMC	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PHI	Primary HIV infection
PI	Protease inhibitor
Pol	Polymerase
PR	Protease
R5-virus	HIV variant using CCR5 coreceptor
Rev	Regulator of virion expression
RNA	Ribonucleic acid
RT	Reverse transcriptase
SIV	Simian immunodeficiency virus
UDPS	Ultra-deep pyrosequencing
V3	Variable loop 3
Vif	Virion infectivity factor
Vpr	Viral protein R
Vpu	Viral protein U
X4-virus	HIV variant using CXCR4 coreceptor

CONTENTS

1	Aims.....	1
2	The human immunodeficiency virus.....	3
2.1	The beginning of the HIV pandemic.....	3
2.1.1	Origin of HIV	3
2.1.2	Global spread	4
2.2	HIV virology	5
2.2.1	Structure and genome	5
2.2.2	Replication	6
2.3	HIV genetic variation	9
2.3.1	Sources of genetic variation.....	9
2.3.2	Genetic variants of HIV.....	9
2.3.3	Methods to study HIV genetic variation	10
2.4	HIV infection	12
2.4.1	HIV-1 transmission	12
2.4.2	HIV-1 infection and pathogenesis	12
2.4.3	Immune responses against HIV-1	14
2.4.4	Coreceptor use.....	14
2.4.5	Tropism testing	16
2.5	Antiretorviral treatment and resistance	18
2.5.1	Antiretroviral treatment.....	18
2.5.2	Monitoring of treatment.....	20
2.5.3	Treatment failure.....	20
2.5.4	Drug resistance	21
2.5.5	Minority drug resistance	24
2.5.6	Transmitted drug resistance	24
2.6	Deep sequencing of HIV-1	25
2.6.1	Next generation sequencing technologies.....	25
2.6.2	Ultra-deep pyrosequencing (UDPS)	27
2.6.3	UDPS data analyses.....	28
3	Materials and Methods	29
3.1	Patient material.....	29
3.2	Methodologies.....	30
3.2.1	UDPS library preparation.....	30
3.2.2	Data filtering	31
3.2.3	Diversity calculations.....	32
3.2.4	Coreceptor use and phylogenetic analyses	33
3.2.5	Ethical considerations	33
4	Results and Discussion	34
4.1	Performance of UDPS	34
4.1.1	Experimental setup.....	34
4.1.2	UDPS evaluation.....	36
4.2	Dynamics of HIV-1 quasispecies.....	38
4.2.1	Minority variants	38
4.2.2	Virologic failure due to drug resistance.....	39

4.2.3	Treatment interruption and drug resistance	40
4.2.4	Coreceptor switch.....	40
4.2.5	Primary HIV infection	40
4.3	Methods to reduce UDPS error frequency	41
4.3.1	UDPS errors and data cleaning approach	41
4.3.2	Reducing errors by molecule-specific tags.....	42
5	Conclusions and future perspectives.....	46
6	Acknowledgements	50
7	References	52

1 AIMS

The specific aims of my thesis were:

- Paper I** To investigate, by ultra-deep pyrosequencing (UDPS), the presence of drug resistance mutations in treatment naïve HIV-1 infected patients and the dynamics of drug resistance development and reversion during treatment initiation and discontinuation.
- Paper II** To evaluate the quality and reproducibility of the UDPS technology in analysis of HIV-1 *pol* gene variation.
- Paper III** To investigate if CXCR4-using virus is present as a minority species already during primary HIV-1 infection in patients whose virus later switches to CXCR4 use.
- Paper IV** To investigate the characteristics and source of errors introduced by UDPS and to develop methods to reduce the error frequency.

2 THE HUMAN IMMUNODEFICIENCY VIRUS

2.1 THE BEGINNING OF THE HIV PANDEMIC

The origin of the human immunodeficiency virus (HIV) virus has been traced to the simian immunodeficiency viruses (SIV), found in African apes and monkeys [1-3]. HIV-1 was introduced to humans through several cross-species transmissions that are estimated to have occurred during the first part of twentieth century in West Central Africa [3-5] but it was only about 30 years ago the recognition and identification of the virus began. In 1981 opportunistic diseases, such as *Pneumocystis carinii* pneumonia and Kaposi's sarcoma, along with immune suppression was reported in young, previously healthy homosexual men in New York City and California [6, 7]. Additional opportunistic complications were soon described, including mycobacterial infections, toxoplasmosis, invasive fungal infections, and non-Hodgkin's lymphoma. The disease was given the name acquired immunodeficiency syndrome (AIDS) [8] but the cause of the disease remained unknown for two more years. The first clear evidence that AIDS was caused by an infectious agent came when a child who received a blood transfusion died of AIDS related opportunistic infections [9]. In 1983, the French researchers Dr Luc Montagnier and Dr Françoise Barre-Sinoussi isolated HIV [10] and in 2008 they received the Nobel Prize for their finding. Since the discovery of HIV, extensive research has shed light on one of the fastest evolving organisms on earth [11]. The ability to rapidly diversify allows HIV to evade the host's immune system [12], alter its cell tropism, and develop resistance to antiretroviral drugs [13].

2.1.1 Origin of HIV

SIVs are known to naturally infect approximately 40 different species of Old World monkeys and apes in sub-Saharan Africa [14]. The zoonotic transmission events of some of these SIVs have resulted in different forms of HIV (type or group) (Figure 1). The transmissions from West Central African chimpanzees (*Pan troglodytes troglodytes*) and from sooty mangabeys (*Cercocebus atys atys*) have been established as HIV type 1 (HIV-1) and HIV type 2 (HIV-2), respectively [1]. The time to the most common recent ancestor (tMRCA) and the origin for HIV-1 and HIV-2 have been estimated using phylogenetic analyses and sequence data with known sampling dates. The result from these studies suggest that the tMRCA for HIV-1 and HIV-2 dates back to 1910 [4] and 1940 [15], respectively.

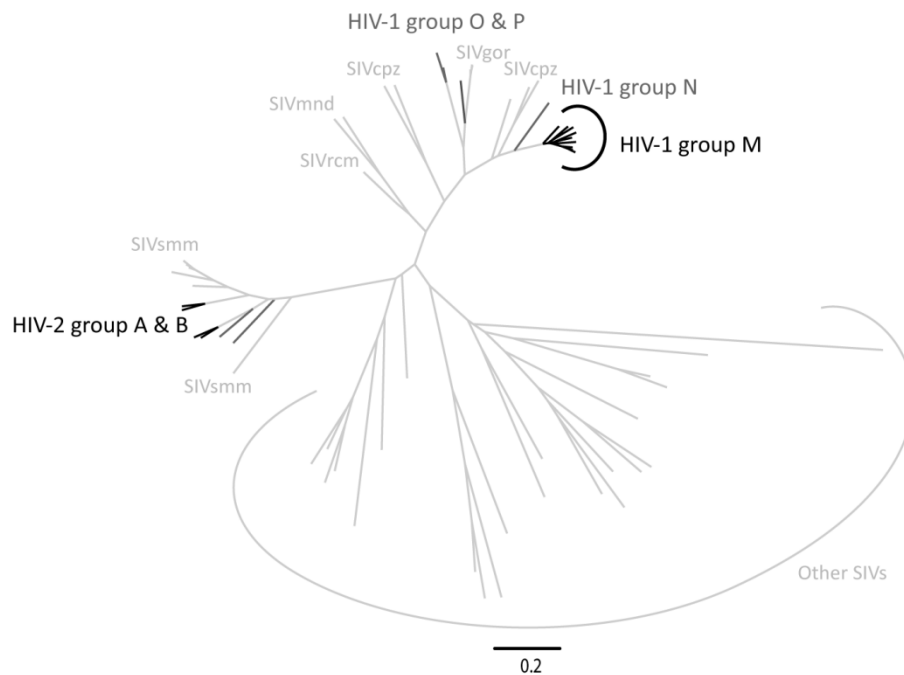


Figure 1. Evolutionary history of the primate lentiviruses. Maximum likelihood tree of the viral *pol* gene. (Kindly provided by Helena Skar).

Interestingly, it has been shown that SIV has been present in African primates for more than 32 000 years [16]. Thus, transmission of SIV to humans has been possible also in the past, however why only the transmissions occurring about 100 years ago where successful remains unknown. This may be explained by social and behavioral changes such as migrations, urbanization [4, 5] and colonization, together with war and health programs [17].

2.1.2 Global spread

Since 1981, more than 60 million people have been infected with HIV-1, and more than 20 million have died from AIDS related disease. The HIV-1 virus has spread to all continents but the most affected part of the world is sub-Saharan Africa, where 22.9 million people live with HIV today [18] (Figure 2). In the western world, HIV-1 infections are more common among populations at higher risk, such as men who have sex with men (MSM), intravenous drug users and immigrants [19]. According to UNAIDS, the global spread of HIV appears to have peaked in 1997, however still the number of people living with HIV is increasing. It could be the refection of combined effects of continued high rates of HIV transmission and the beneficial impact of antiretroviral treatment [20].

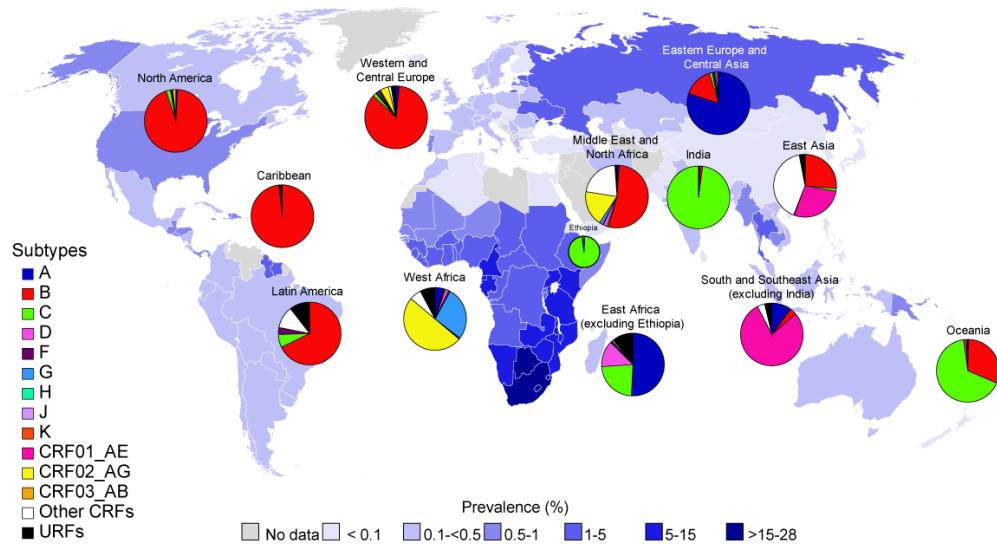


Figure 2. The global HIV-1 subtype distribution. Reprinted with permission from [21].

2.2 HIV VIROLOGY

2.2.1 Structure and genome

HIV belongs to the Lentivirus genus of the *Retroviridae* family. Retroviruses are enveloped viruses that contain two copies of positive-sense single-stranded RNA molecules, which are non-covalently linked at the 5'-end. The HIV virus contains a conical nucleocapsid that surrounds the viral nucleic acid as well as the viral enzymes, reverse transcriptase (RT), protease (PR) and integrase (IN), which are required for the early replication events. The envelope consists mainly of host cell lipid bilayer membrane together with viral trimeric glycoprotein gp41 covalently linked to the external trimeric gp120 (Figure 3).

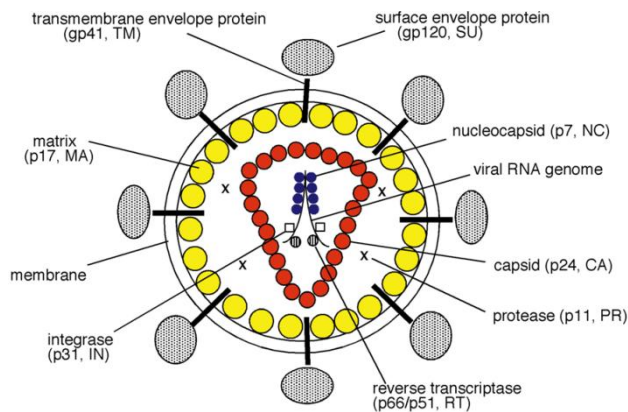


Figure 3. Schematic structure of the HIV particle. Reprinted with permission from [22].

The HIV genome is approximately 10,000 nucleotides in length. Like other retroviruses, it has three major structural genes: group-specific antigens (*gag*), polymerase (*pol*) and envelope (*env*) (Figure 4). The HIV-1 *gag* gene encodes the polyprotein precursor p55, which is processed into p24 (capsid), p17 (matrix), p7 (nucleocapsid), and p6 by the viral protease. HIV-1 *pol* encodes the viral enzymes

PR, RT and IN. The *gag* and *pol* genes are produced as Gag or Gag-Pol precursor polyproteins that are cleaved by the viral PR into the functional proteins. The *env* gene encodes the viral polyprotein gp160/gp140 that is cleaved into the external glycoprotein gp120 and the transmembrane protein gp41, which are important for viral attachment to the host CD4 receptor and fusion with the host membrane. In addition, HIV-1 has two regulatory genes: *tat*, *rev* and four accessory genes: *vif*, *vpr*, *vpu* and *nef*. These genes are important for the viral lifecycle of HIV-1, summarized in Table 1.

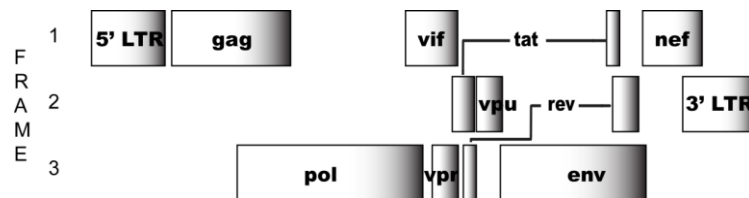


Figure 4. The genomic organization of HIV-1.

Table 1. Regulatory and accessory proteins

Gene/ Protein	Time of expression	Present in virion	Function
tat/Tat	Early	No	Transactivator of HIV gene expression. It binds to the TAR RNA element to facilitate initiation and elongation of viral transcription.
rev/Rev	Early	No	Regulation of viral expression. Permits unspliced mRNA to exit the nucleus into the cytoplasm.
nef/Nef	Early	Yes	Negative regulatory factor. Down regulates CD4 and MHC class I and class II.
vpr/Vpr	Late	Yes	Viral protein r. Causes G2/M arrest, thus preventing cell division. Vpr is also involved in the import of the pre-integration complex (PIC) into the nucleus.
vpu/Vpu	Late	No	Viral protein u. Promotes degradation of CD4 in ER and enhances release of virions from the plasma membrane.
vif/Vif	Late	Yes	The virion infectivity factor. It inhibits the antiviral APOBEC protein and thus G-to-A hypermutations.

2.2.2 Replication

The virus particle infects cells by gp120 binding to the primary receptor, the CD4 molecule, on the target cell (Figure 5). The CD4 receptor is present on CD4⁺ T-lymphocytes, macrophages, monocytes, dendritic cells and microglia cells in the central nervous system [23, 24]. After binding, gp120 undergoes conformational changes that enable interaction with the coreceptor, most often CCR5 or CXCR4

[25], see part 2.4.4 below. Binding to the coreceptor brings the virion in close contact to the cellular membrane, allowing a part of gp41 to penetrate the cell membrane. This penetration mediates fusion of the virus envelope with the cell membrane and release of the viral nucleocapsid into the cytoplasm. The RT enzyme present in the nucleocapsid converts ssRNA into dsDNA, inside the partially opened capsid. Reverse transcription is primed with a human transfer RNA (tRNA) that is bound to the viral RNA inside the virion. Although HIV carries two strands of RNA only one DNA molecule is formed per virion [26]. One of the two copies of ssRNA is thought to act as an alternative template for reverse transcription if the RT encounters a nick or break during reverse transcription [27]. The ribonuclease H (RNase H) domain of the HIV polymerase degrades the viral RNA after reverse transcription. During the transcription, parts of the long terminal repeats (LTRs) are generated in the 5'- and 3'-ends of the genome. These LTRs are important for the integration process and for transcription and translation. The pre-integration complex (PIC) is subsequently transported into the nucleus, where the HIV integrase catalyzes the integration of the viral dsDNA into the host genome. The integration preferably takes place into active and thus open regions of the human genome [28], but integration can also take place in resting cells [29-32]. Once integrated the viral DNA is referred to as a provirus that remains permanently associated with the host genome. The provirus can remain in a latent state (and be passed on to daughter cells by cell division) or be activated and transcribed into viral mRNA by the host RNA polymerase II.

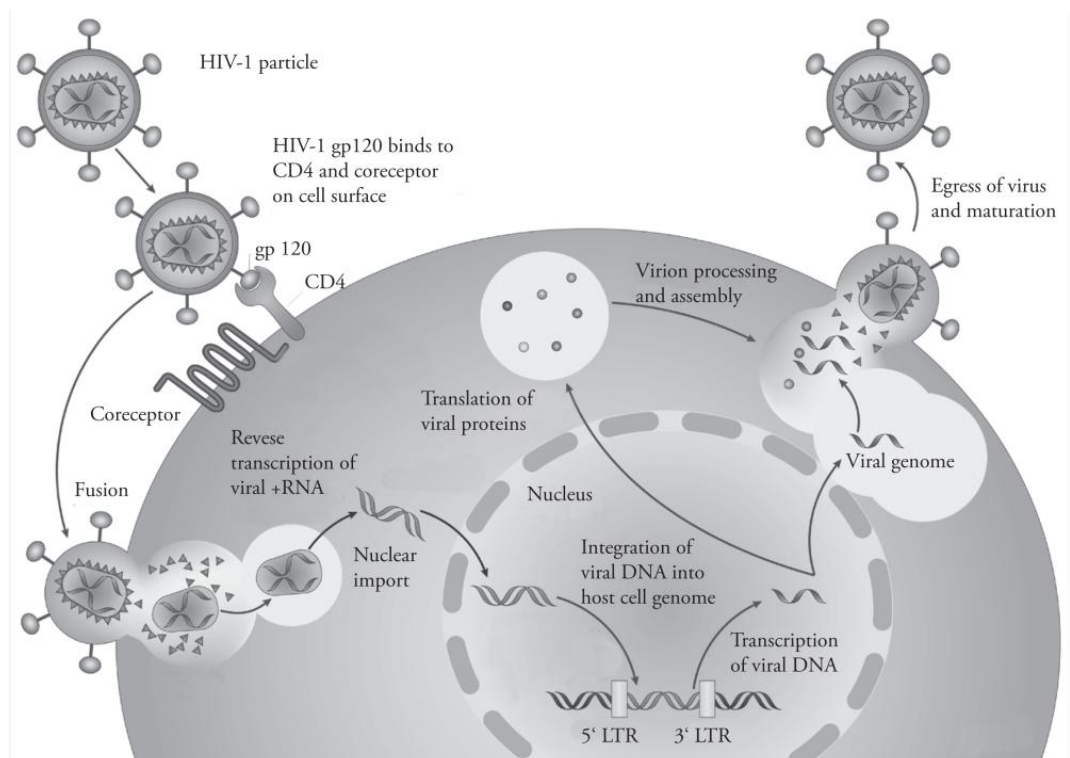


Figure 5. The replication cycle of HIV-1. HIV-1 enters target cell through interactions with CD4 and a coreceptor. The RNA is reverse transcribed and inserted into the host cell genome. Transcription and translation is performed by the cellular machinery. New viral particles are assembled at the plasma membrane. After budding the viral protease cleaves the Gag-Pol precursor polyproteins into functional proteins, which generates a mature infectious virus particle. Adapted from [33].

A single promoter in the 5' LTR region mediates transcription of the HIV-1 genome. The LTR region contains binding sites for several transcription factors. Expression from the 5' LTR generates a 9-kb primary transcript that has the potential to encode all nine HIV genes. The primary transcript can be spliced into several mRNA species or packed without further modification into new virion particles. The early fully spliced transcripts encode the Nef, Tat and Rev proteins in a Rev-independent manner. The Tat protein interacts with the transactivation response element (TAR), located downstream of the LTR region, to greatly increase the levels of transcription of viral RNAs. Thus the Tat protein plays a key role in the activation and maintenance of high levels of transcription from proviral DNA [34, 35]. The Rev protein binds to the rev responsive element (RRE) in the env region of the HIV mRNA and functions as a carrier of the unspliced or partially spliced RNA from the nucleus to the cytoplasm.

The late transcription involves expression of the longer gag, gag-pol, env, vif, vpr and vpu mRNAs, which are unspliced or incompletely spliced and therefore require Rev in order to be transported to the cytoplasm. All mRNAs are translated in the cytoplasm near the endoplasmatic reticulum (ER) by the normal cellular transcription machinery. The envelope protein (gp160) is processed in the ER and the Golgi complex, where it is cleaved by cellular proteases into the surface proteins gp41 and gp120 and heavily glycosylated. Finally, gp41 and gp120 are transported to the plasma membrane of the cell.

The assembly of new HIV particles begins at the plasma membrane. Two HIV ssRNA molecules together with Gag (p55) poly-protein, Gag-Pol (p160), Vif and Vpr associates as the virion begins budding from the host cell. Vpu has not been detected in virus particles [36]. The accessory protein Vif counteracts the antiviral activity of apolipoprotein B mRNA editing enzyme-catalytic polypeptide-like 3G (APOBEC3G), by facilitating its degradation and thus prevents its virion encapsidation [37, 38]. APOBEC3G contributes to an innate resistance to retrovirus infection by deamination of cytidine (C) to uridine (U) in minus strand reverse transcripts, a process that results in guanosine (G) to adenosine (A) mutation of the plus strand DNA [39]. Hypermutation usually results in the production of replication-incompetent virus due to the introduction of stop codons. The A-rich genome of HIV is believed to partly be due to the activity of APOBEC3G.

The immature HIV-1 particle buds from the cell, but has poor ability to fuse with targets cells because of an interaction between Gag and a cytoplasmic tail of gp41[40]. After budding, the viral PR cleaves the Gag structural polyprotein precursor into matrix (MA), capsid (CA), nucleocapsid (NC) and p6 proteins [41]. These proteins form the mature nucleocapsid and matrix, making the virus particle infectious [42]. The viral protease also cleaves the Gag-Pol polyprotein into the viral enzymes: PR, IN and RT.

2.3 HIV GENETIC VARIATION

2.3.1 Sources of genetic variation

HIV is one of the fastest evolving organisms known. Due to the fast evolutionary rate, the virus evades the host immune system and has the capacity to develop resistance to antiretroviral drugs during suboptimal treatment. There are at least six mechanisms that contribute to the high genetic variation of HIV:

- I) The error-prone RT enzyme generates on average 0.1-0.3 mutations per genome and replication cycle [43-45] and it is considered to account for most of the point mutations seen in HIV-1. These mutations remain uncorrected since RT lacks proofreading activity.
- II) The RT enzyme switches between the two ssRNA strands during reverse transcription and it has been estimated that such recombination events occur between 2 to 30 times per replication cycle [46-48]. For recombination to contribute to evolution the template switching needs to result in a novel genetic variant, which only happens when the two strands in the infecting virus are different. This is referred to as effective recombination rate. The effective recombination rate has been estimated to be $1.4 \pm 0.6 \times 10^{-5}$ recombinations per site and generation, which assumes a probability of coinfection of about 10% [49].
- III) HIV-1 establishes a lifelong infection with continuous replication and high viral production rate. In untreated patients approximately 10^{10} new virions are produced every day. These virions have an average life-span of 2-3 days [50-52].
- IV) The immune system exerts a high selective pressure on the viral population [53, 54]. Therefore, immune-escape variants often have a survival advantage and become subject to positive Darwinian selection.
- V) The cellular RNA polymerase II, which transcribes the integrated proviral DNA into mRNA, is also error-prone due to a lack of proofreading activity.
- VI) Finally, the effect of the cellular enzyme APOBEC3G, may (if not completely inhibited by Vif) result in excess G-to-A mutations. However, viruses with hypermutations resulting in stop codons are nonviable and do not contribute to evolution.

The reason why certain mutations survive to the next generation and eventually becomes fixed in the population is dependent on a combination of selective pressures, fitness costs and chance events.

2.3.2 Genetic variants of HIV

The high genetic variability of HIV-1 has a direct effect on within-patient evolution. In a patient HIV-1 variants can differ by more than 5%, which is a greater genetic distance than between the human and the mouse genomes. In addition, the high genetic variability of HIV has given rise to series of phylogenetically defined groups and clades (subtypes), seen on the population level. To date, HIV-1 is divided into three groups, group M (main), O (outlier) and

N (non-M non-O) (Figure 1), which probably represent independent transmission events from chimpanzees. In 2006, SIV was discovered among the western lowland gorillas (*Gorilla gorilla gorilla*) and the virus was genetically linked to HIV-1 group O and the not yet formally approved fourth group, group P. However, chimpanzees are thought to be the original reservoir for SIV_{gor} as well [55]. HIV-1 group M has successfully spread to all continents on earth and is further divided into nine subtypes (A, B, C, D, F, G, H, J and K) [56]. In addition, more than 50 circulating recombinant forms (CRFs) have been recognized so far (<http://www.hiv.lanl.gov>). The CRFs are recombinant viruses which have been formed in patients infected with more than one subtype. The HIV subtypes and CRFs have spread unevenly around the world. The highest diversity of HIV remains in the western part of Africa and despite the potential for divergent viruses to spread only a few subtypes have successfully expanded. About 90% of the epidemic comprises of four subtypes (A, B, C and D) and two circulating recombinant forms (CRFs) (CRF01_AE and CRF02_AG) [57]. HIV-2 is divided into group A and B, which both are endemic in West Africa. In contrast to HIV-1, HIV-2 have had limited spread to other parts of the world. Founder effects, whereby a single chance introduction into a naïve population causes massive spread probably account for most of the current geographic distribution of HIV genetic variants, but human genetics, behavioral factors and possibly viral fitness differences may also have contributed.

2.3.3 Methods to study HIV genetic variation

Genetic variation and the relationship between sequences can be visualized in a phylogenetic tree. It can be used to study evolutionary relatedness of different organisms or relationship between strains of the same organism. Due to the fast evolution of the HIV virus, it is possible to use phylogenetic trees for detailed evolutionary and epidemiological studies. The branching-pattern of the tree is called the topology and the length of the branches describes their genetic distances, which is related to their evolutionary time. The sequences represented by the tips are called taxa. There are several different methods to infer a phylogenetic tree. The four main methods are: Neighbor joining, Parsimony, Maximum likelihood and Bayesian inference (reviewed in [58]). These methods are briefly described in table 2.

To infer a phylogenetic tree the model of sequence evolution (substitution model) first needs to be selected. One of the simplest models assumes that all nucleotides occur with the same frequency and that all point mutations occur at the same rate. However, since the evolutionary process often is more complex, this method most often underestimates genetic changes. For instance, transitions are usually more common than transversions. Several different substitution models have been proposed to more realistically describe sequence evolution by accounting for unbalanced base composition and mutation rates. The most complex substitution model is GTR (general time-reversible) model, in which each pair of nucleotide substitutions has different rates, i.e. it assumes a time reversible symmetric substitution matrix in which A is substituted by T with the same rate as T substitutes to A. Mutation rates usually also differ across sites of the genome.

There exist several methods to account for these rate variations. The most commonly used adds a gamma-distributed rate parameter (G) to the substitutions model. Furthermore, information about invariant sites (I), can also be added to the model. Thus, GTR+G+I represents a complex model that often recapitulates HIV-1 evolution fairly realistically. To accurately infer a phylogenetic tree the best-fit substitution model, G and I should be estimated from the data. In general, the simplest model that adequately explains the data should be used. Several programs can be used to obtain the best-fit model, such as jModelTest [59] or FindModel (<http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html>).

Table 2. Description of phylogenetic methods

Method	Description	Advantages	Disadvantages	Examples of Software
Neighbor joining (NJ)	Creates a pair-wise distance matrix describing the evolutionary distance between sequences, which is used to construct the tree.	Fast; Works well on closely related sequences.	Information is lost by compressing sequences into distances. Does not explore many tree options.	MEGA [60] PAUP*
Parsimony	Chooses between trees to find the one with the least number of mutations that describes the data.	Relatively fast; Works well on closely related sequences.	Can perform poorly if the distances between sequences varies.	PAUP* MEGA [60]
Maximum likelihood (ML)	Uses a statistical model to find the tree that has the highest likelihood of producing the observed data given the assumptions.	More accurate than NJ and Parsimony on distantly related sequences and/or rapidly evolving organisms. Explores a large tree space.	High computational burden.	PAUP* Phylip PhyML [61]
Bayesian inference	Based on a statistical model. Simultaneously estimates trees and uncertainty for every branch.	More accurate than NJ and Parsimony on distant sequences and/or rapidly evolving organisms. Explores a large tree space and outputs a collection of trees that fit the data.	High computational burden; The prior distributions for parameters needs to be specified.	MrBayes Beast

There are different ways of assessing confidence of the branches in the tree. The traditional method is called bootstrap analysis, where the original alignment is randomly re-sampled with replacement to produce pseudo-replicate data-sets. New trees are inferred on these datasets and offer measurements of which part of the tree has higher or lower support. The main drawback of bootstrapping is the

computational burden, since the original analysis is repeated for each pseudo-replicate dataset [58], i.e. at least 100 and often to 1000 times. There are other alternatives of assessing confidence of the tree topology such as the approximate likelihood-ratio test (aLTR) where the significance of a branch is tested based on the null hypothesis corresponding to the assumption that the inferred branch has length 0 [62]. Moreover, Bayesian methods have assessment of confidence naturally implemented in the methodology. It is important to note that a tree is the best attempt to explain the data given the model, which is not necessarily the same as the evolutionary history.

2.4 HIV INFECTION

2.4.1 HIV-1 transmission

Globally, sexual transmission accounts for approximately 80% of all HIV infections, where heterosexual intercourse accounts for the majority of transmissions [18]. The risk of transmission of HIV-1 is 0.01-0.23% after a single heterosexual exposure. Higher viral load and genital ulceration are important determinants of HIV-1 transmission per coital act [63]. Transmissions can also occur by transfer of contaminated blood through needle stick injuries, the sharing of contaminated needles between intravenous drug users or from mother to child during pregnancy, *in utero*, during delivery or postnatal through breastfeeding.

Infectiousness is correlated to the viral load and therefore is especially high during primary infection when the viral load temporarily is very high. Antiretroviral treatment dramatically lowers viral load and several studies have demonstrating its potential for prevention of HIV transmission [64-66].

Several studies have shown that the HIV-1 diversity is low during primary HIV infection (PHI) [67-71] and that most HIV-1 infections probably are established by one or a few virus particles [70, 72, 73]. However it is still uncertain if more virus particles actually are transmitted but only one or a few viruses grows out. Transmission bottlenecks have been seen not only in mucosal transmission, but also in infections through intravenous drug use [74]. The diversity has been shown to gradually increase during the course of infection in the absence of treatment [75-77]. However, the diversity has been suggested to decrease in late infection [78]. Furthermore, a reduction in evolutionary rate has been shown to coincide with disease progression [79].

2.4.2 HIV-1 infection and pathogenesis

The course of HIV-1 infection can be divided into three stages: the acute stage, the chronic stage and AIDS. Immediately after exposure and transmission, the virus cannot be detected in plasma. This so called eclipse phase generally lasts 7 to 21 days [70, 80, 81]. The reason for this is probably that HIV-1 replicates in the mucosa, submucosa and draining lymphoreticular tissues (such as gut-associated lymphoid tissue). Once HIV-1 reaches a concentration of 20 copies per milliliter in

plasma it can be detected by quantitative clinical assays. Studies implicate that CD4⁺ T-lymphocytes and Langerhans' cells are the first targets of the virus [82, 83] and that monocyte-derived macrophages are generally poor targets as compared with CD4⁺ T-lymphocytes [73]. Studies have shown that HIV-1 rapidly replicates, first in the gut-associated lymphoid tissue and then systemically [84], resulting in a rapid increasing plasma viral loads to 10⁷ – 10⁸ RNA copies per milliliter at peak viremia, which occurs approximately 25 days after infection [81]. In the gut-associated lymphoid tissue the phenotype of the most productively infected cells appears to be the resting CD4⁺ T-lymphocytes that lack activation markers and expressing low levels of the chemokine receptor CCR5 [85]. Instead, many of these cells express the $\alpha_4\beta_7$ integrin receptor [86]. Regardless of the route of transmission an irreversible destruction of reservoirs of helper T-lymphocytes, especially in the gut-associated lymphoid tissue, is seen which has implications on the pathogenesis of HIV infection.

The acute phase, which also is called primary HIV infection (PHI) is characterized by high viral loads and the sequential appearance of viral markers and antibodies in the blood. Approximately, 50% of patients infected with HIV will develop symptoms of acute HIV infection. Early HIV infection can be divided into stages, called Fiebig stages [81], based on the detection of HIV-1 antigens and HIV-1-specific antibodies in diagnostic assays (Figure 6). The acute infection phase is divided into five Fiebig stages (I-V) and the early chronic HIV infection is defined as Fiebig stage VI, occurring at approximately 100 days following infection.

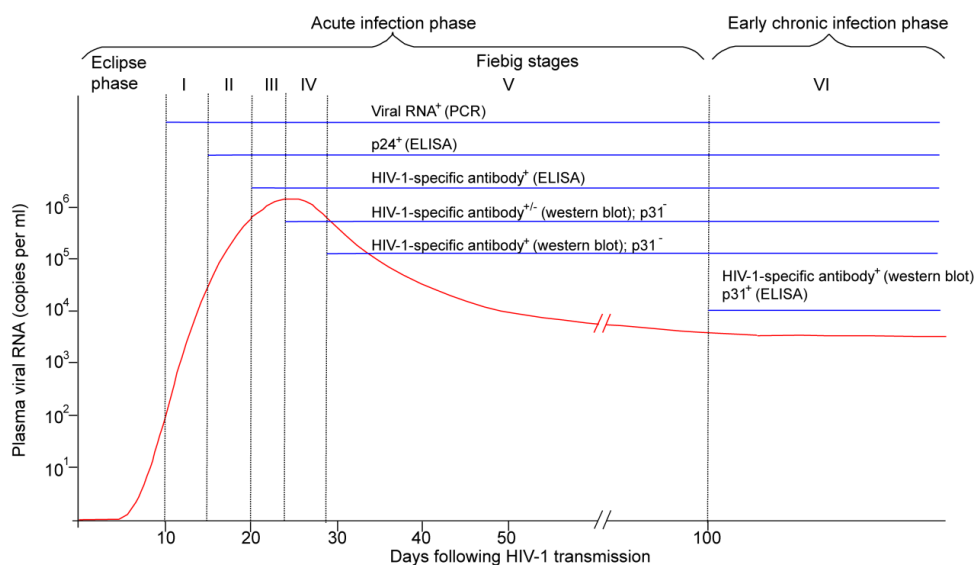


Figure 6. Early HIV-1 infection. The first weeks after infection are divided into Fiebig stages that are defined by a stepwise gain in positivity for the detection of HIV-1 antigens and HIV-1-specific antibodies in diagnostic assays. Adapted from [87].

The chronic phase is characterized by the establishment of a viral setpoint and partial restoration of CD4⁺ T-lymphocyte levels. The setpoint has been shown to be predictive of disease progression in HIV-1 infection as individuals with high plasma HIV-1 RNA levels progress more rapidly to AIDS than those with low levels [88]. The average viral setpoint is around 30,000 HIV-1 RNA copies per milliliter plasma in HIV-1 infected patients, which is a level that has been suggested to

maximize the transmission potential by mathematical modeling [89]. In the absence of treatment the average time to onset of AIDS is around 10 years [90]. AIDS results from long-term (chronic) HIV infection, where the immune system has been exhausted by the constant battle of the infection, and specific opportunistic infections or malignancies are diagnosed. In addition, in the US an absolute CD4 cell count of less than 200 cells/ μ l also constitutes an AIDS criterion.

Some patients remain asymptomatic for more than 10-15 years and are called long-term non-progressors. Many of these patients have certain genetic traits, especially certain HLA types (e.g. HLA B57 and HLA B27), which have been associated with delayed HIV-1 escape and a decreased rate of disease progression. In contrast, there are other HLA types that are associated with an increased rate of disease progression (e.g. B35-Px), where patients progress to AIDS within 2-3 years [91]. The delta-32 deletion in the CCR5 gene (CCR5 Δ 32) is another genetic trait that either causes high resistance to infection or delay disease progression, when present homozygous or heterozygous, respectively [92, 93].

2.4.3 Immune responses against HIV-1

The first line of defense in response to HIV-1 infection is the innate immune system, followed by the development of adaptive immune responses. The initial decline of plasma viral load after peak viremia during acute infection is thought to be due to mainly CD8⁺ T-lymphocyte-mediated killing of productively infected cells [94, 95]. Thus, HIV-1 specific T cell responses develop before seroconversion and just before the peak viremia is reached. However, HIV usually rapidly escapes these first T cell responses, indicating that many targeted epitopes are readily changeable. T cell responses targeting more slowly evolving or conserved epitopes develops later. These later T cell responses may be important in lowering and maintaining the viral set-point [94].

Antibodies directed against HIV-1 have been seen to arise within eight days of infection. These first antibodies forms immune complexes and are not likely to impact on the control of acute phase viremia, however early escape from neutralizing antibodies has been reported [96-98]. No association has been seen between specific antibody responses and natural control of HIV-1 viremia during chronic infection. Before onset of late HIV-1 infection, the humoral immune system constantly changes specificity to target new HIV-1 variants. Even though some of these antibodies may be neutralizing they lag behind, rarely targeting the contemporary viruses [54, 98-100].

2.4.4 Coreceptor use

The entry of HIV-1 into target cells is dependent on the binding of the viral envelope glycoprotein to its receptor CD4 and a coreceptor, most often C-C chemokine receptor type 5 (CCR5) or C-X-C motif receptor 4 (CXCR4) (Figure 7) [101-105]. Several additional co-receptors have been identified *in vitro*, but only the CCR5 and CXCR4 appear to have a major role in HIV-1 attachment *in vivo* [106]. Some viruses can use both CCR5 and CXCR4 coreceptor (R5X4-viruses). R5X4 viruses are also called dual/mixed to signify that some assays do not

distinguish between viruses consisting of truly dual tropic clones and those with mixtures of R5 and X4 clones [77]. Although CXCR4-using viruses (X4-virus) have been shown to be transmissible [107], the majority of infections are established by CCR5 using viruses (R5-virus). During transmission only one or a few viral particles establish the infection even though the inoculum most probably contains more virus variants. One theory suggests that this transmission bottleneck is the result of selection acting on the envelope gene, which favors CCR5-using viruses during transmission and/or establishment of successful infection [67, 68, 108]. Indeed, individuals homozygous for the delta-32 deletion in the CCR5-gene seem to be protected against HIV-1 infection. However other studies have argued that there is no conclusive evidence to support that CXCR4 using variants are less transmissible [77, 109]. In about 50-70% of HIV-1 infected patients the viral population switches to include X4-virus later in infection [77, 110-112]. The emergence of X4-virus is temporally associated with accelerated CD4⁺ T-lymphocyte decline and progression to AIDS [77, 110, 111, 113, 114]. This pathogenic difference was known already in the late 1980's, before the coreceptors were identified. At that time, the replicative capacity of HIV-1 variants in peripheral blood mononuclear cells (PBMC) was referred to as rapid/high or slow/low [115, 116], and the capacity of inducing syncytia in PBMC or MT-2 cells was identified (referred to as syncytium inducing (SI) and non-syncytium inducing (NSI)) [117-120]. About a decade later the coreceptors were identified [101-105] and since then there has been intense research to try to understand the complex mechanisms behind coreceptor switch. However, it is still not known if the emergence of X4-viruses is a cause and/or a consequence of immunodeficiency [121].

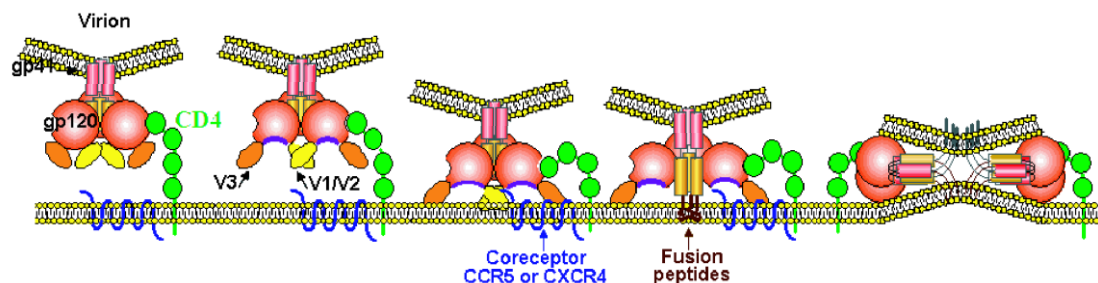


Figure 7. Schematic illustration of HIV-1 entry into target cell. Gp120 binds to CD4, which induces conformational changes in gp120 and exposure of the coreceptor binding site. Conformational changes in gp41 allows insertion of the fusion peptide into the host cell membrane. During the final step the six-helix bundle is formed, which brings the viral membrane and the host membrane together and allows fusion. (Kindly provided by Dr R.W Doms).

The CXCR4 receptor is mainly expressed on naïve CD4⁺ T-lymphocytes, whereas memory CD4⁺ T-lymphocytes mainly expresses CCR5. Since the CCR5 receptor is also expressed on macrophages it was first believed that R5-viruses were macrophage-tropic (M-tropic) and that X4-viruses were T-lymphocyte tropic (T-tropic). However, it has been shown that monocyte-derived macrophages are generally poor targets for primary HIV-1 isolates as compared to CD4⁺ T-lymphocytes [73]. Thus, both R5 and X4 virus are mainly CD4⁺ T-lymphocyte tropic.

The CCR5 and CXCR4 coreceptors are chemokine receptors that belong to the seven transmembrane spanning G-protein-coupled receptors that are involved in signal transduction [122, 123]. The natural ligands (chemokines) for CCR5 are RANTES (regulated on activation, normal T-cell expressed, and secreted), macrophage inflammatory protein (MIP) -1 α and MIP- 1 β . For CXCR4 the natural ligand is stromal cell-derived factor 1 (SDF-1). These chemokines exhibit suppressive effect on HIV-1 by down regulating coreceptor expression and by competitive binding [124, 125]. In addition, individuals with the CCR5 Δ 32 are protected against HIV-1 infection with R5-viruses (which establish most new infections). Thus, the development of the drug maraviroc, which blocks the CCR5 coreceptor, was a quite logic step to inhibit HIV-1 replication (see section 2.5.1 about antiretroviral treatment).

The viral envelope proteins gp41 and gp120 are glycosylated in the endoplasmic reticulum where they are produced. The surface of gp120 consists of five constant regions (C1-C5) and five variable regions (V1-V5). The principal determinant of coreceptor use is the variable loop 3 (V3) [126], but parts of V1/V2, V4 and C4 have also been shown to impact coreceptor use [127-129]. Both the CD4 binding site and the co-receptor binding site are partly masked by the hypervariable V1/V2 loop structure. Attachment between gp120 and the CD4 molecule displaces the V1/V2 loop and V3, creating the coreceptor binding site [130, 131]. The V3 loop is a 35 amino acid long loop structure held together by a disulphide bond between the cysteins at position 1 and 35. It has been shown that a few amino acid changes in V3 can change the coreceptor use from CCR5 to CXCR4 [132, 133]. Electrostatic interactions have a major role in coreceptor binding [134]. Thus, the presence of basic amino acids (lysine or arginine) at positions 11 and 25 is associated with CXCR4 use [133], whereas acidic or uncharged amino acid in position 11, 25 or 28/29, resulting in a low V3 charge is associated with CCR5 use [133, 135, 136]. Moreover, the V3 charge increased with time in R5 populations from patients with virus populations that switch coreceptor use, while it remains unchanged or decreased in non-switch populations[137]. Glycosylation of the envelope spikes have been shown to be important for the folding of gp120 upon binding as well as determinants of the coreceptor usage of HIV-1 [138, 139]. In addition, this host derived glycan-shield hinders efficient antibody binding, thus impairing immune recognition. During the course of infection, the glycosylation sites in the HIV-1 envelope gene continuously changes leading to an evolving glycan-shield [99]. In contrast, during primary infection the level of glycosylation of the envelope spikes has been shown to be lower [140].

2.4.5 Tropism testing

Coreceptor tropism refers to the ability of HIV-1 to enter CD4 cells by the CCR5, CXCR4 or both coreceptors (dual tropism) [141]. Viral tropism can be assessed by genotypic or phenotypic approaches. The first widely used phenotypic method was the MT-2 assay. In this assay patient-derived cells or established isolates are co-cultured with MT-2 cells, which express CXCR4 coreceptor but not the CCR5 coreceptor [142]. X4 and dual-tropic R5X4 viruses are capable of infecting the MT-

2 cells which results in the formation of large syncytia that are visible by light microscopy. The viral replication can also be assessed by detection of viral antigen in culture supernatant. The main drawback of this assay is that no negative control is used. Thus, if a virus does not grow in the MT-2 assay, it might be due to technical difficulties that prevent infection or because it is a R5 virus. A second drawback is that it is labor intensive and requires viral culturing in a BSL3 facility and a third drawback is that it typically tests PBMC virus rather than plasma virus.

Today, recombinant phenotypic assays are available, such as the Trofile assay [143]. In this assay, the entire patient-derived *env* gene is amplified directly from plasma by PCR and inserted into an expression vector. This vector and a replication-defective proviral vector containing a luciferase reporter gene are co-transfected in a HEK293 cell line to produce a pseudovirus population, which is subsequently used to infect U87 cell lines expressing either CXCR4 or CCR5 receptor. Infection is assessed by quantifiable light emission. Co-receptor antagonists are added as additional controls. The reliability of this assay depends mainly on the sensitivity and accuracy of the cDNA synthesis and PCR and proportion of HIV-1 population amplified. The assay can be used with plasma HIV-1 RNA loads greater than 1000 copies per milliliter and X4 variants that comprise 0.3% of the population can be detected with 100% sensitivity [144]. The test can be done on both RNA and DNA but in Europe the commercial test is available only for plasma RNA. Other similar assays exist, such as the Toulouse Tropism Test [145], however insufficient data exist to assess the reliability of this assay for samples with low viral loads.

Genotypic tropism testing is based on sequencing of the V3 region of the HIV-1 *env* gene directly from patients plasma samples [146-148]. Either population based sequencing (Sanger sequencing) or ultra-deep pyrosequencing (UDPS) approaches have been used for both viral RNA and DNA. The phenotypes of the sequences are predicted by bioinformatic interpretation techniques, such as 11/25 charge rule, the position-specific scoring matrix (PSSM) and geno2pheno (G2P). Briefly, the 11/25 charge rule is the simplest algorithm, which takes only the charge of the amino acids at key position 11 and 25 in the V3 loop into account. In comparative studies, only a moderate correlation with results from the original Trofile assay was reported.

PSSM is a more advanced method, where the sequences' likelihood of being derived from an X4 virus for every possible amino acid at every individual position is calculated [149]. There are two matrices available for determining scores in subtype B: i) X4R5, which is calculated using sequences with known coreceptor phenotype as indicated by growth on indicator cells expressing CD4 and either CCR5 or CXCR4. ii) SINIS, is calculated using sequences producing syncytium on the MT2 cell line. In either case, the input sequences are compared and aligned to sequences of known coreceptor use (e.g. X4). The better the fit, the higher PSSM score and the higher the score the higher likelihood that the sequence fragment has X4 properties. Sequences with values above -2.88 are considered X4, whereas sequences with scores below -6.96 are considered R5.

Sequences with intermediate scores cannot be predicted using this method. This method can be accessed online: <http://indra.mullins.microbiol.washington.edu/webpssm/>

Another advanced prediction method is G2P_[coreceptor] [150]. This method is based on a statistical learning method called a support vector machine which is trained with a set of nucleotide sequences that corresponds to R5, dual/mixed tropism or X4 phenotypes. Nucleotide sequences are used as input. The result of interpretation is given as a false positive rate (FPR), which is defined as the probability of falsely classify an R5 virus as X4. The European guidelines on the clinical management of HIV-1 tropism testing recommend that a FPR of 5.75% should be used [151]. This method can be accessed online at: <http://coreceptor.bioinf.mpi-inf.mpg.de/index.php>.

None of the available genotypic prediction methods take additional regions of *env*, outside the V3 loop into account. This means that sites that might be important in e.g. V1/V2 are missed. PSSM and G2P have been evaluated in several studies and clinical trials. In a study published by Harrigan *et al.*, these methods were compared with the original Trofile assay. The sensitivities were 56 and 63% and specificities were 90 and 91% for the two assays, respectively [152]. Indeed, it is important to note the concordance between phenotypic and genotypic methods is not perfect [107, 153, 154] and quite commonly the genotypic prediction tools falsely predict R5 variants as X4 variants [107]. These rates of false positives might not be a problem when screening prior to maraviroc use, especially if other treatment options exists, however when searching for rare cases of X4 variants (e.g. X4/X4R5 transmission) in UDPS studies, the predictions needs to be interpreted with caution [109].

2.5 ANTIRETORVIRAL TREATMENT AND RESISTANCE

2.5.1 Antiretroviral treatment

Without the use of antiretroviral treatment, almost all HIV-1 infected patients would die from AIDS. In 1987, the first drug for HIV treatment was approved. It was zidovudine (AZT), a nucleoside reverse transcriptase inhibitor (NRTI) that interferes with HIV replication by competitively inhibiting the reverse transcriptase enzyme, resulting in chain termination during viral DNA synthesis [155, 156]. In the 1990's additional NRTIs, non-nucleoside reverse transcriptase inhibitors (NNRTIs) became available (Table 2). However, HIV quickly developed resistance to these drugs since they were used in mono- or dual therapy regimens. It was not until 1996 when drugs from at least two different drug classes, NRTIs, NNRTIs and protease inhibitors (PIs), were used in triple combination, called highly active antiretroviral therapy (HAART) or combination antiretroviral therapy (cART), that the morbidity and mortality of HIV-1 infected patients were greatly reduced [157-159]. Successful HAART dramatically suppresses viral replication and reduces the plasma viral load to below limits of detection of the most sensitive clinical assays (<20 RNA copies/mL). However, despite HAART low levels of free virions can be found in the plasma. Whether this residual viremia

represents ongoing cycles of replication [160] or simply the release of virus from stable reservoirs [161-166] is controversial. Since no viral evolution has been confirmed the later explanation is more likely [167].

Table 2. Antiretroviral drugs approved by Food and Drug Association (FDA) and European Medicines Agency (EMA)

Drug	Approved FDA/EMA	Drug mechanism
NRTIs		
abacavir (ABC)	1998/1999	NRTIs are first activated into the 5'-triphosphate form by host enzymes. The active NRTIs compete with RT's natural substrates (dNTPs) and when incorporated they function as chain terminators, lacking a 3'-hydroxyl group necessary for elongation [168].
didanosine (ddI)	1991 *	
emtricitabine (FTC)	2003/2003	
lamivudine (3TC)	1995/1996	
stavudine (d4T)	1994/1996 [#]	
tenofovir (TDF)	2001/2002	
zalcitabine (ddC)	1992 *	
zidovudine (AZT)	1987/1987	
NNRTIs		
delavirdine (DLV)	1997/-	NNRTIs inhibit DNA polymerization by binding a small hydrophobic pocket near the RT active site, which induces a conformation change of the substrate-binding site and reduces polymerase activity [169].
efavirenz (EFV)	1998/1999	
etravirine (ETR)	2008/2008	
nevirapine (NVP)	1996/1998	
rilpivirine	2011/2011	
PIs		
Atazanavir (ATV)	2003/2004	Most PIs are peptidic or peptidomimetic compounds designed as analogs of the cleavage sites found within the Gag and Gag-Pol precursor proteins. Some PIs are transition state analogues that resemble the transition state of a substrate molecule in the PI catalyzed reaction. PIs have poor oral bioavailability and most PIs are thus co-administrated with low dose ritonavir, an HIV-1 protease inhibitor that inhibits the 3A4 isozyme of cytochrome P ₄₅₀ (CYP 3A4), which is responsible for the metabolism of most of these drugs [170].
Darunavir	2006/2008	
Fosamprenavir (fAMP)	2003/2004	
Indinavir (IDV)	1996/1996	
Lopinavir (LPV)	2000/2001	
Nelfinavir (NFV)	1997/1998 ^a	
Saquinavir (SQV)	1995/1996	
Tipranavir (TPV)	2005/2005	
Fusion inhibitors		
Enfuvirtide (T-20)	2003/2003	Enfuvirtide is a peptide drug selected from chemically synthesized peptides derived from various regions of gp41 [171]. The peptide sequence binds to gp41, preventing the formation of the hairpin structure (six-helix bundle) and consequently, the fusion [172]. It is sensitive to proteolytic digestion and needs to be administered by injection.
Entry inhibitors		
Maraviroc (MVC)	2007/2007	Maraviroc is a noncompetitive, specific, slowly reversible CCR5 coreceptor antagonist that selectively binds to the human chemokine receptor CCR5 present on the host cell membrane. Binding alters the conformation of the receptor and prevents interaction with the V3 loop to CCR5, and the subsequent membrane fusion [173, 174]. HIV-1 tropism test such as the Trofile assay is recommended before use.
Integrase inhibitors		
Raltegravir (RAL)	2007/2007	The integrase inhibitor binds to the specific complex between integrase and the viral DNA and thus selectively targets the strand transfer reaction of the integration reaction [175, 176].

*withdrawn from market by manufacturer. [#]not recommended by Swedish guidelines due to side effects. ^anot recommended by Swedish guidelines due to low antiviral activity.

HAART reduces the rate of sexual transmission, which has additional public health benefits [64]. However, HAART alone cannot eliminate HIV-1 infection since virus is hidden in the latent reservoirs [177]. Viral reservoirs have been found in a variety of cell types including CD4⁺ T-lymphocytes [177, 178], dendritic cells (DCs) [179-181] and macrophages [178]. Resting memory CD4⁺ T-lymphocytes decay very slowly during HAART, with an average half-life of 44 months, indicating that under current treatments it will take over 60 years to deplete this reservoir [182]. Therefore, HIV infected patients cannot be cured from the infection by current treatment options.

In the beginning of 21st century three new drug classes were introduced: fusion inhibitors, entry inhibitors and integrase inhibitors (Table 2). In the western world, where HAART is affordable, HIV infection has turned from a deadly infection into a chronic life-long disease. However, in the developing world, especially in low and middle- income countries about 53% of eligible HIV infected patients have not yet started HAART [20].

2.5.2 Monitoring of treatment

Disease progression is monitored by CD4 cell counts, HIV-1 plasma RNA levels and clinical symptoms. In untreated patients, CD4 cell counts is the most important marker, while treated patients are primarily monitored by measuring HIV-1 RNA levels in plasma. HIV-1 treatment guidelines in the US and European Union recommend the initiation of HAART when the CD4 cells in peripheral blood decline to 350 cells per μ L. The recommended first line HAART regimen, consists of two NRTIs and either a NNRTI or a PI, and are quite similar in the US [183] and in Europe [184]. The Swedish guidelines are summarized in Table 3 [185]. Combination regimens consisting of raltegravir and two NRTIs are also recommended as initial regimens in the US and European guidelines, but not in the Swedish guidelines.

Table 3

Preferred 1 st line regimen*	Drugs
NNRTI-based	efavirenz + abacavir/lamivudine or tenofovir/emtricitabin
PI-based	atazanavir/r + abacavir/lamivudine or tenofovir/emtricitabin durunavir/r + abacavir/lamivudine or tenofovir/emtricitabin

* Swedish guidelines for antiretroviral therapy (www.smittskyddsinstitutet.se/rav) [185].

2.5.3 Treatment failure

Treatment failure of HAART naïve patients can be caused by several factors, including poor adherence, pharmacologic factors such as drug-drug interactions that impair absorption or accelerate clearance, host factors (e.g. low CD4⁺ cell count at start of therapy), transmitted drug resistance or drug resistance development during treatment [186]. There are three types of treatment failure: virologic failure, immunologic failure and clinical progression. Virologic failure is when the viral load rebounds or does not decrease sufficiently despite HAART. Immunologic failure is when the CD4⁺ T-cell counts do not increase despite

HAART. Finally, clinical progression is when symptoms of HIV disease occur despite HAART.

During suboptimal treatment (e.g. mono or dual drug combinations as well as insufficient adherence to HAART) selection of pre-existing variants with reduced susceptibility or development of *de novo* resistance mutations can occur [187]. Thus, for an HIV treatment to be successful, patients need to be committed and adherent to reduce the possibility of drug resistance development.

2.5.4 Drug resistance

The drug resistant variants usually have reduced fitness compared to wild-type virus. This is especially true for viruses with single primary resistance mutations. In contrast, additional mutations, which may evolve over time during continued drug selective pressure, may be compensatory, thus restoring fitness to near wild-type levels. The rate of development of drug resistance depends on patient adherence to treatment, the genetic barrier (see below), host genetics, and fitness of the drug resistant variant [188, 189]. The emergence of drug resistance has been shown to be associated with an increased mortality among patients first starting HAART [190]. Thus, the clinical management of HIV-1 infection is important to reduce the risk of treatment failure. Genotypic HIV-1 resistance testing is an important tool for clinical management HIV-1 infection. Population based sequencing of *pol* gene (including PR, RT and when required also IN) is generally generated by in-house methods or by commercial assays such as ViroSeq from Abbott. The sequences can be used for online prediction at Stanford (<http://hivdb.stanford.edu>) and National Agency for AIDS Research (www.medpocket.com).

The terminology used in the field of drug resistance classification can be confusing and no universal system exists. Thus, drug resistance mutations are classified differently by different systems, which also change over time. Here, I use the definition from the latest update from the International AIDS society - USA [191], where PI mutations are classified into major and minor mutations depending on when they are selected. Major mutations are defined as those selected first in the presence of the drug or those substantially reducing drug susceptibility. Minor mutations generally emerge later and do not by themselves have a strong effect on phenotype. However, minor mutations may improve fitness of viruses containing major drug resistance mutations. NRTI and NNRTI mutations are not classified into major and minor mutations by the IAS-US system, instead the first mutations that arise are referred to as primary mutations. Furthermore, HIV drugs can be divided into low or high genetic barrier to resistance depending on the number of mutations needed and the fitness cost of these mutations to the virus. Most NRTIs and NNRTIs are generally considered to have low genetic barrier whereas PIs are considered high genetic barrier to resistance (Figure 8).

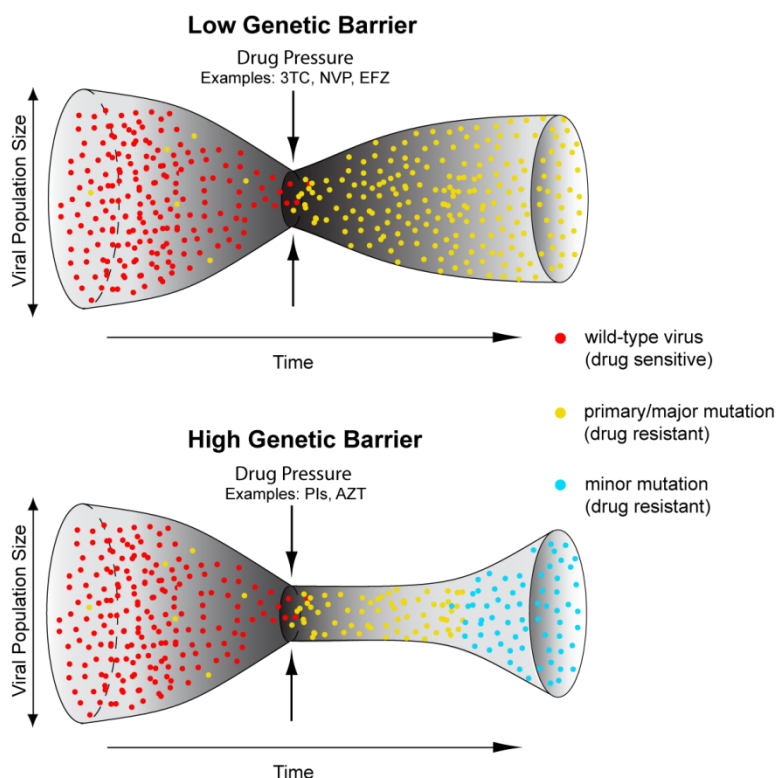


Figure 8. The development of drug resistance to low and high genetic barrier drugs. Low genetic barrier drugs: Selection of pre-existing minority variants with primary/major mutations. High genetic barrier drugs: Selection of primary/major mutations followed by *de novo* evolution of minor mutations. Red: Wild-type virus. Yellow: Primary/major mutations are the first mutations that arise and often have a strong effect on resistance and a fitness cost and to e.g. lamivudine (3TC), zidovudine (AZT) and PIs. Blue: Minor mutations usually have little or no effect on resistance, but restore fitness. Adopted from [21].

In many cases the drug resistance mutations alters binding site for the nucleoside or NRTI, thus preventing incorporation of the drug into the nascent chain. Mutations associated with this mechanism include the M184V/I and K65R. The M184V/I mutation can emerge with 3TC or FTC therapy [192, 193]. For AZT however, the mutations do not prevent the binding and incorporation of AZT triphosphate into the growing chain, but rather seem to activate a reverse reaction by which the AZT nucleotide is removed from the chain, subsequently permitting normal elongation [194]. These mutations are called thymidine analog mutations (TAMs) and they promote pyrophosphorolysis and are involved in the excision of AZT and d4T [195]. TAM amino acid changes in HIV-1 RT include two distinct pathways: the TAM1 pathway (M41L, L210W, T215Y, and occasionally D67N) and the TAM2 pathway (D67N, K70R, T215F and 219E/Q) [187, 196].

NNRTI resistance generally result from single amino acid substitutions such as K103N and Y181C [197, 198]. Most NNRTI resistance mutations cause some level of cross-resistance among different NNRTIs. In contrast to NRTI resistance mutations, which often are associated with reduced fitness, single nucleotide changes associated with NNRTI resistance can result in high-level resistance with only a slight loss of fitness [199, 200]. The low genetic barrier, minimal impact on fitness and the slow reversion of NNRTI mutations in patients in the absence of

drug contribute to transmission and stability of NNRTI-resistant HIV-1 in the population [201]. In HIV-1 group O and HIV-2, the 181C are commonly found as wild-type variants [202], thus NNRTIs are ineffective against these viruses. The mutations associated with NRTI and NNRTI drug resistance are summarised in Figure 9.

NRTI drug resistance mutations

M	K	D	K	L	Y	M	L	T	K
41	65	67	70	74	115	184	210	215	219
L	R	N	R	V	F	V	W	Y	Q
			E			I		F	E
d4T AZT	ABC ddI FTC 3TC d4T TDF	d4T AZT	d4T TDF AZT	ABC ddI	ABC	ABC FTC 3TC	d4T AZT	d4T AZT	d4T AZT

NNRTI drug resistance mutations

	L	K	K	V	V	Y	Y	G	P
	100	101	103	106	108	181	188	190	225
	I	P	N	M	I	C	C	S	H
		E		A		I	L	A	
		H		I		V	H		
	EFV NVP	EFV ETR NVP	EFV NVP	EFV ETR NVP	EFV NVP	EFV ETR NVP	EFV NVP	EFV ETR NVP	EFV

Figure 9. Mutations associated with NRTI and NNRTI drug resistance. All mutations associated with ETR are not shown. Adapted from [191].

Resistance to PIs was initially expected to be low, because of the vital role of this enzyme in the life cycle of HIV-1 and its relatively small size. However, due to the plasticity of the protease, drug resistance can develop also to PIs. PI resistance requires a stepwise accumulation of major and minor (compensatory) mutations [203] and since several mutations are needed the genetic barrier to resistance is higher for PIs compared with NRTIs and NNRTIs (Figure 8). Unlike NNRTI resistance mutations, major drug-resistant PI mutations are rarely observed in the viral populations in protease naïve individuals. Finally, most PIs share similar chemical structure and cross-resistance is commonly observed.

Resistance to integrase inhibitors is almost always associated with mutations within the integrase active site [176]. These mutations have deleterious effect on the enzymatic function of the enzyme and have high fitness cost to the virus. However, development of compensatory mutations will somewhat restore fitness [201]. Mutations that confer resistance to the fusion inhibitors T-20 also results in reduced replicative fitness, probably because mutations that reduce T-20 binding also reduce the efficiency of six-helix bundle formation which is essential for viral fusion [172]. The resistance profile towards the entry inhibitor maraviroc is different from the other HIV-1 drugs because it binds to a host receptor. Resistance to maraviroc can either confer tropism switching, increase affinity to

the coreceptor or use of the inhibitor-bound receptor for entry [201]. However, evaluations of coreceptor use in clinical trials have shown that a coreceptor switch to CXCR4 occurs in patients with pre-existing CXCR4-using virus and does not seem to be a favoured resistance pathway [204].

2.5.5 Minority drug resistance

Due to the high mutation rate of HIV-1, it has been predicted that drug resistance mutations spontaneously develop every day in untreated patients [205]. Thus pre-existing drug resistance to low genetic barrier drug such as lamivudine and nevirapine is likely to be present also in patients that have never received ART and not been infected by resistant virus. The clinical relevance of these mutations is not yet fully understood.

Recently, several reports indicate that low-frequency or minority drug resistance HIV-1 variants have clinical significance. It has been shown that these variants adversely affect the response to HAART thus increasing the risk of virological failure, especially if mutations associated with resistance to NNRTIs are present [206-212]. However, in other studies no association between minority variants and virologic failure has been found [213, 214]. Available data suggest that pre-existing minority resistance mutations may have greater impact on the clinical utility of drugs with low genetic barrier than those with high genetic barrier [215, 216]. These minority resistance mutations are thought to mainly represent unrecognised transmitted resistance rather than spontaneous resistance mutations due to replication infidelity. In contrast, minority X4 or R5X4 viruses may arise spontaneously as a first step towards coreceptor switch and may lead to therapy failure to regimens containing maraviroc [217]. It is not known at which frequency of X4 virus the maraviroc treatment is still active. Prior to maraviroc use the European guidelines [184] and the Swedish guidelines [185] currently favor genotypic testing for determining coreceptor use, while the US guidelines favor the phenotypic Trofile assay [183].

2.5.6 Transmitted drug resistance

Transmitted drug resistance (TDR) is a clinical and epidemiological problem because it may reduce the efficacy of antiretroviral treatment. The prevalence of TDR ranges from 5.6% in Sweden to 14.6% in the U.S. [218-222]. The level of drug resistance in patients infected with a drug resistant founder virus is usually low with only one or a few drug resistance mutations, which is somewhat difficult to explain because most treated patients have virus with more resistance mutations. This could be due to that virus with lower number of drug resistance mutations have higher transmission fitness. Another reason could be that the transmitted drug resistant virus had more mutations, but that these may have fully or partially reverted to wild-type after transmission due to their fitness costs. It has been shown that some mutations quickly revert to wild-type (e.g. M184V) while others appear to persist for many years (e.g. M41L) [223-226]. These differences most likely reflect the combined effect on replicative capacity of the mutations, the reversion possibilities and the presence of compensatory mutations [227]. Partially reverted mutations may be present as minority mutations, which might

be detected with deep sequencing techniques (see section 2.6 on UDPS), even though they remain hidden with routine resistance tests that are based on population Sanger sequencing. In addition, HIV-1 infects long-lived memory CD4⁺ T-lymphocytes and it is therefore likely that the drug resistant founder variant(s) will be stored in these archives along with a representation of all other variants that evolve over time in the individual. Furthermore, if the selection pressure changes, these archived variants can reappear, which has been shown for both wild-type variants and resistant variants [228-231]. The reappearance of resistant virus has been associated with treatment failure [230, 231].

2.6 DEEP SEQUENCING OF HIV-1

2.6.1 Next generation sequencing technologies

Over the past six years new sequence technologies have been developed, which have revolutionized genomic science. These technologies are referred to as next generation sequencing (NGS) technologies. The population based Sanger method [232] is considered as a 'first generation' technology, and dominated the sequencing field for about two decades prior to the development of NGS [233]. The high throughput of the NGS technologies, generating million of sequence reads in a short time, makes them suitable for shotgun sequencing of whole genomes, such as human, bacteria and plants. In shotgun sequencing, short sequence reads are generated from fragmentized DNA. The reads are aligned to a known reference sequence or assembled *de novo*. Sequencing platforms that generates short reads and huge amount of data are suitable for such projects (e.g. SOLiD and HiSeq). For projects where smaller genomes or fragments are analyzed, usually platforms such as 454 sequencing, Ion Torrent or MiSeq are also suitable. Most NGS platforms offers the possibility of deep sequencing, where for instance rare HIV-1 drug resistance mutations in a background of wild-type variants can be analyzed. Due to the length of the reads, 454 sequencing has an advantage for deep sequencing projects, referred to as ultra-deep pyrosequencing (UDPS). In Table 4, the most commonly used technologies and their pros and cons are described.

The NGS technologies described above are constantly improving, generating longer and increased throughput. In the last couple of years new NGS technologies have been developed, such as Heliscope from Helicos Biosciences and PacBio from Pacific Biosciences. The Heliscope and PacBio promises to resolve true single molecule sequencing. It is too early to say if these techniques will get a marketplace. This will depend on their performance in this highly competitive field. It is also likely that the several NGS platforms will coexist in the marketplace, due to specific features that makes some more advantageous for specific application compared to others [233].

Table 4. NGS technologies

Sequence technology	Description	Read length (bp)	Number of reads per run	Pros	Cons
Sanger sequencing	Chain termination method with dye-labeled dideoxynucleotides. Capillary electrophoresis and fluorescence detection that provides four-color plots to reveal the DNA sequence.	1000	96	Population based sequencing. Low error rate.	No high throughput. Can only detect polymorphism $\geq 20\%$.
454 GS FLX titanium and 454 junior/ Roche	The DNA templates are clonally amplified on beads (one template/bead) in an emulsion based PCR [234]. The beads are transferred to a PicoTiterPlate. Sequencing chemistry by pyrosequencing, where the intensity of the bioluminescence is recorded and the DNA sequence is revealed.	200 ¹ 500 ^{2,3}	1 million 70,000 ³	Longer reads improve linkage analysis.	High errors frequency in homopolymer regions.
MiSeq and HiSeq/ Illumina	The DNA templates are clonally amplified on a solid plate generating clusters [235]. Sequencing by reversible terminators in a cyclic method that comprises nucleotide incorporation, fluorescence imaging and cleavage.	2 x 100 ⁴	600 million - 3 billion	Currently dominates the NGS market for shut-gun sequencing.	Time consuming library preparation, long run times.
SOLiD/ Life Technologies	The DNA templates are clonally amplified on beads (one template/bead) in an emulsion based PCR [234]. Sequencing chemistry by sequencing by ligation of cleavable probes. Fluorescence imaging to determine the ligated probe.	60	1 - 3 billion	2 + 1 base encoding provides inherent error correction.	Long run times, short reads.
IonTorrent /Life Technologies	The DNA templates are clonally amplified on beads (one template/bead) in an emulsion based PCR [234]. Sequencing based on sequencing by synthesis technology where release of H ⁺ is measured as a result of an incorporated nucleotide.	100 200	1.2 -to 660 million	Scalable Less labor intensive compared to 454 and Illumina.	High errors frequency in homopolymer regions (similar to 454).

¹ 454 GS FLX, ²454 GS FLX titanium, ³454 GS Junior System. ⁴Paired-end sequencing. Adapted from [233].

2.6.2 Ultra-deep pyrosequencing (UDPS)

One application of NGS is deep sequencing, referred to as ultra-deep pyrosequencing (UDPS), which is performed on the 454 platform. UDPS allows identification of rare genetic variants, which are not detectable by population based Sanger sequencing [207, 236-238]. The depth of the UDPS analysis is primarily determined by the number of templates that can be successfully extracted and amplified from the starting material and by the error rate of PCR and UDPS.

Figure 10 schematically describes the different steps in the UDPS technology. First the DNA library is prepared, by amplifying the region of interest. The A and B adaptor are either synthesised as a part of the PCR primers or subsequently ligated to the amplicons. These adaptor sequences are required for binding of the amplicons to the beads during the emulsion based PCR (em-PCR) and enables bidirectional pyrosequencing. The DNA amplicons are diluted so that statistically only one molecule will bind to each bead. The beads are separated into individual water droplets in an emulsion, where each water droplet is its' own PCR reactor, which enables each DNA molecule to be clonally-amplified on the bead. The beads with bound DNA are enriched and distributed on a titanium coated PicoTiterPlate (the titanium coated plates became available in 2008 and increased read length and improved quality by reducing crosstalks between adjacent wells), which is a fiber optic slide, containing millions of small wells, where only one bead will fit in each well. Enzyme beads (containing sulfurylase and luciferase) are layered into the wells of the PicoTiterPlate. The sequencing reaction is based on a sequencing-by-synthesis method called pyrosequencing [239]. The sequencing reagents (containing buffers and nucleotides) are flowed across the wells of the plate. The nucleotides are added sequentially in a fixed order across the PicoTiterPlate during the sequencing run. During the nucleotide flow, millions of copies of DNA bound to each of the beads are sequenced in parallel. When a nucleotide complementary to the template strand is added into a well, the polymerase extends the existing DNA strand by adding nucleotide(s). Addition of one (or more) nucleotide(s) generates a light signal that is recorded by the CCD camera in the instrument. The signal strength is proportional to the number of nucleotides incorporated. Data are stored in standard flowgram format (SFF) files for downstream analyses.

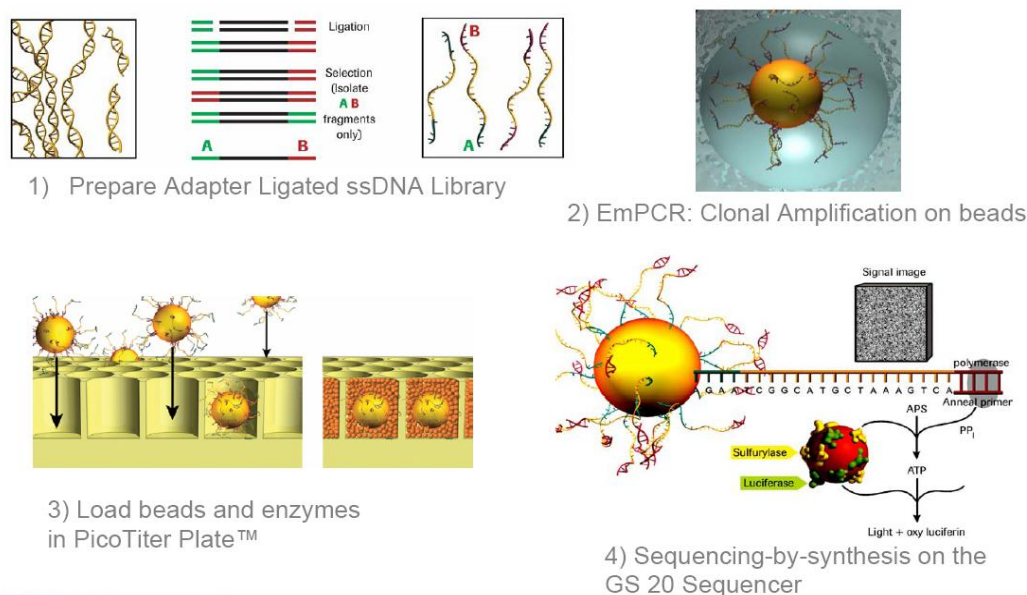


Figure 10. Overview of the 454 sequencing system.

2.6.3 UDPS data analyses

The data produced from UDPS place substantial demands on data handling and quality control. The UDPS technology has higher error rate compared to Sanger sequencing, which means that it may be difficult to distinguish rare, but genuine, variants from artifacts generated in the cDNA synthesis, PCR or sequencing reaction. Homopolymeric stretches (of three or more of the same nucleotide) are particularly problematic due to difficulties in interpretation of the chemiluminiscent signal. Researchers are using different bioinformatic approaches to identify, remove or correct these sequencing artifacts, by in-house cleaning strategies [228, 240-242] or bioinformatic programs such as ShoRAH [243] and Segminator [244]. Which strategy to use is dependent on the specific data set and factor such as read-length and if assembly tools needs to be used.

3 MATERIALS AND METHODS

3.1 PATIENT MATERIAL

All patient samples used in the studies were retrieved from the HIV biobank at the Swedish Institute for Communicable Diseases. The biobank was established in 1985 and contains more than 35,000 clinical samples (serum, plasma, PBMC and/or HIV isolates). The patient plasma samples used in **Paper I**, **Paper II** and **Paper III**, were collected between 1989 and 2003 and have been stored at -70°C or -20°C. Some samples had been freeze-thawed prior to our analyses.

In **Paper I**, six HIV-1 subtype-B-infected individuals who experienced virological treatment failure were selected and a total of 40 plasma samples from these patients were analyzed. All patients, except one, had later undergone treatment interruption. The treatment histories of the patients differed, but all started therapy before the HAART era and all had received regimens that contained 3TC, AZT and d4T. Five of six patients were sampled before any treatment was initiated. All patients were sampled two to six times during treatment, i.e. the first available sample after treatment initiation and additional samples obtained during therapy failure. Finally, five patients were sampled following treatment interruption.

In **Paper II**, four plasma samples used in **Paper I** were selected. Samples A, B, C and D in **Paper II** are the same as samples 6.4, 2.5, 4.5 and 3.5 from **Paper I**, respectively. The first number corresponds to the patient number and the second number corresponds to the sample in the series. Sample A and B were used to study repeatability, effects of sequencing direction and influence of primer related selective amplification. Plasma samples C and D were used to generate two molecular clones (clone 1 and clone 2) for studies on UDPS sensitivity and *in vitro* PCR recombination. These two clones were chosen on the basis of sequence dissimilarity with the aim to maximize the number of informative sites in the *pol* amplicon of interest.

In **Paper III**, three HIV-1 subtype-B-infected patients with HIV-1 populations that switched coreceptor use to include CXCR4 during infection were selected and four to nine plasma samples from each patient were investigated. Information about coreceptor use was available in the database that is linked to the biobank and was based on MT-2 phenotypic tests performed on PBMC isolates at the time of sampling. For two patients, the first samples were obtained during early PHI. Thus, when the first study sample was obtained, patient 1 and patient 2 had a negative HIV-1 antibody test and positive HIV-1 antigen and RNA tests, which is consistent with a classification into Fiebig stage II. Patient 3 was classified into Fiebig stage IV-V based on a positive antibody test on ELISA and incomplete Western blot pattern (see section 2.4.2 on HIV-1 infection and pathogenesis).

In **Paper IV**, patient material was not used, however sequences from control samples (clones) from **Paper I** were studied in more detail. The characteristics and reproducibility of sequencing errors was investigated in a SG3Δenv plasmid clone datasets that had been generated in three separate UDPS runs.

3.2 METHODOLOGIES

3.2.1 UDPS library preparation

The depth of UDPS analyses depends on the error frequency of the method (see section 3.2.2 below) and the number of input DNA templates. To obtain a high recovery of templates it is important to use sensitive and robust methods for RNA extraction, cDNA synthesis and PCR amplification. In **Paper I**, substantial efforts were invested in evaluating and comparing different approaches for RNA extraction, cDNA synthesis and PCR amplification, with the aim to maximize the number of plasma HIV-1 RNA molecules that were extracted, reverse transcribed, PCR amplified and finally subjected to UDPS. The final optimized protocol is presented below in Figure 11 and was used in **all papers**. Each sample was tagged with sample-specific sequence tags during the PCR to enable multiplexed UDPS sequencing. Before UDPS, the PCR amplicons were purified and the DNA concentration and purity was determined using Nanodrop, Qubit and Agilent 2100 bioanalyzer. After these quality controls, the PCR amplicons were pooled and sequenced in both forward and reverse direction on the 454 Life Science platform according to the manufacturer's instructions. The 454 GS FLX, with a read length of about 200 bp, was used in **all papers**. In addition, the 454 GS Junior System was used for some samples in **Paper III**. For detailed information about the specific procedure see materials and methods in **Paper I and Paper III**.

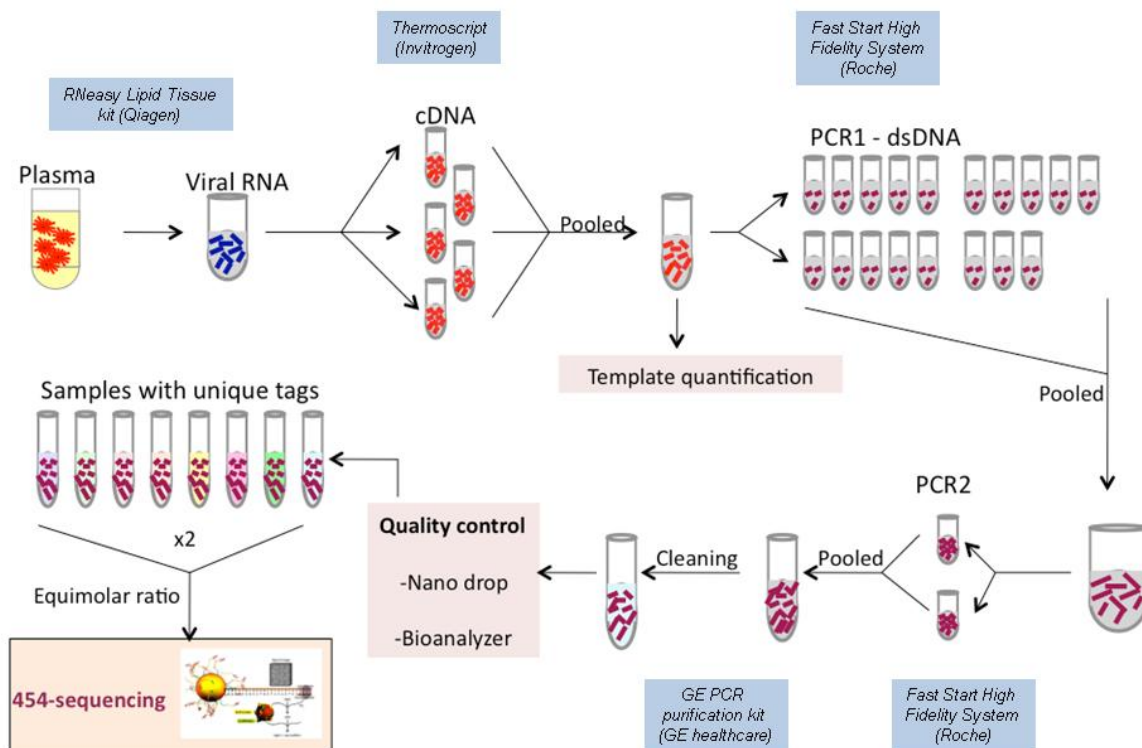


Figure 11. Schematic illustration over the experimental setup used in all papers.

In **Paper I**, **Paper II** and **Paper IV** a fragment of 167 nucleotides in the *pol* gene (positions 3058 to 3226 in HxB2, GenBank accession number K03455) corresponding to amino acids 171 to 224 in RT was amplified (Figure 12). In **Paper III**, a fragment covering the V3 region in the *env* gene (positions 7010 to 7332 in HxB2, GenBank accession number K03455) was amplified.

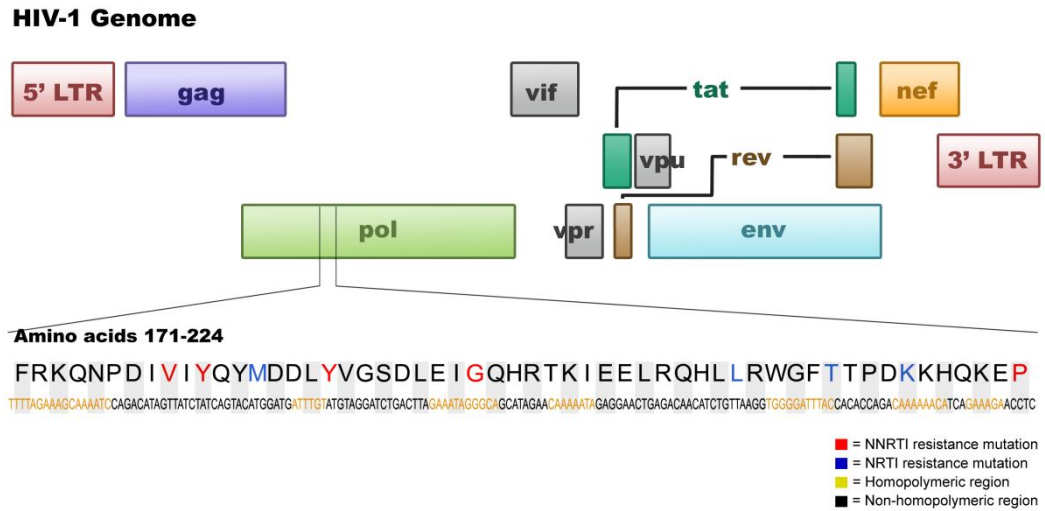


Figure 12. HIV-1 *pol* amplicon. The NRTI and NNRTI drug resistance mutations are shown in blue and red, respectively. The drug resistance positions studied in Paper I are the NRTI mutations M184V/I, L210W, T215Y/F, K219Q/E and the NNRTI mutations Y181C/I/V, Y188C/L/H, G190S/A. (Kindly provided by Johanna Brodin).

3.2.2 Data filtering

Bioinformatic software was written in PERL to manage, clean and analyze the UDPS data. The data cleaning method removed reads with characteristics associated with UDPS sequencing errors. Unlike some other methods we did not attempt to correct errors since we did not want to risk creating sequence variants that did not exist in the original patient sample. A similar data handling and filtering strategy was used in **all papers** and is summarized below.

1. Each sample was identified by their sample-specific sequence tags.
2. Reads with <80% similarity (**Paper I, II and IV**) or <70% similarity (**Paper III**) to the corresponding Sanger sequence were filtered.
3. Reads containing ambiguous bases (Ns) were filtered.
4. Reads that did not cover the entire region of interest (amino acids 180–220 in RT, position 3087 to 3206 in HxB2, GenBank accession number K03455) (**Paper I, II and IV**) were filtered.
5. Remaining reads were imported into the GS amplicon software (Roche, Penzberg, Germany) and aligned.
6. The data were compressed by PERL scripts that identified unique sequence variants in forward and reverse direction (**Paper I, II and IV**) and counted the number of reads per variant. In **Paper III**, forward and reverse reads were combined, due to low number of reads in the reverse direction compared to the forward direction. The tally for each variant was retained as part of the sequence name for further analyses.

7. The alignment was extracted and cut to the region of interest amino acid 180–220 in *pol* (**Paper I**, and **II**) or position 7137 to 7242 in the V3 region (**Paper III**). In **Paper IV**, the entire length of the *pol* reads were retained.
8. Since UDPS errors are known to be concentrated to homopolymeric regions, reads with out-of frame insertions or deletions were removed, while reads with in-frame indels (i.e. ± 3 , 6, 9 nucleotides) were retained.
9. Finally, the alignments were manually inspected and any remaining variants with frameshifts or stop codons were removed.
10. For **Paper I**, **II** and **IV**, the unique variants found in forward and reverse direction were compared and the abundance of the variant was set to the sum of the forward and reverse tallies. However, if frequencies of the forward and reverse reads differed by more than a factor of 10 we made the assumption that a systematic error had occurred during UDPS and adjusted the frequency of the variant to the lower of the two estimates. Finally, variants were discarded from further analyses if the variant was absent in either forward or reverse direction. For **Paper IV**, no more filtering was done beyond this step.
11. In **Paper I** and **Paper II**, both drug resistance analysis and individual variant analysis were done on the datasets using two different cut-off approaches:
 - **Paper I** - Drug resistance analyses: individual average cut-off values (with a 95% confidence interval) were calculated for each drug resistance mutation positions using data from the SG3 Δ env plasmid sequenced on three occasions. These cut-off values, which were adjusted to each sample (since different number of reads were obtained for each sample), were used to evaluate if the frequency of drug resistance were significantly higher than the background error rate at that position (Chi-square test with correction for continuity).
 - **Paper I** and **Paper II** - Individual variant analyses: variants were classified as high-confidence variants or as probable sequencing artifact by using the overall average error rate generated from the three SG3 Δ env plasmid datasets as cut-off. Variants with prevalence significantly higher than the cut-off value was retained and variants below the cut-off were discarded.
12. In **Paper III**, the number of reads from each sample was adjusted to the template molecule availability. Since the number of templates varied between samples due to differences in viral load, high-confidence variants could not be calculated based on cut-off values. Instead variants represented by one read were removed to reduce the dataset before phylogenetic analyses.

3.2.3 Diversity calculations

In **Paper I** and **Paper III**, the pair-wise distances for each sample were calculated using MEGA 4 with the Tamura-Nei model with gamma distributed rates across sites ($\alpha=0.5$). The average genetic distance per sample was calculated using an in-house PERL script that weighted sequence variants according to their abundance.

3.2.4 Coreceptor use and phylogenetic analyses

In **Paper III**, phenotypic coreceptor testing had been determined by the MT-2 assay at the time of sampling. The coreceptor use of each individual V3 sequence was predicted using the bioinformatic algorithms PSSM_{x4/r5} and geno2pheno_[coreceptor]. Variants were predicted to use CXCR4 if the PSSM score was above -2.88 and the geno2pheno FPR was below 5.75%, according to European guidelines [151]. Variants that fulfilled only one of these criteria were considered to have an uncertain coreceptor use, whereas variants that did not fulfill any of the two criteria were considered to use CCR5. The evolutionary relationships were analyzed by maximum likelihood trees constructed using PhyML 3.0 and the best-fit-model of nucleotide substitution identified by jModelTest.

3.2.5 Ethical considerations

For **Paper I**, **Paper II** and **Paper III**, ethics application was approved by Regional Ethical Review board in Stockholm, Sweden (Dnr 52/2008-77). In **Paper IV**, no patient material was used and therefore no ethics application was needed.

4 RESULTS AND DISCUSSION

In this thesis, I have used UDPS to dissect the HIV-1 quasispecies, with the aim of studying HIV-1 evolution during drug resistance development and coreceptor switch, **Paper I** and **Paper III**, respectively. To enable deep sequencing with the UDPS methodology, we carefully optimized all protocols upstream from UDPS (**Paper I**) and developed bioinformatic tools to clean and handle the data (**Paper I** and **Paper IV**). In **Paper II**, we evaluated the performance of our UDPS methodology. Finally, preliminary results from an ongoing project, with the aim to further reduce the error frequency by introducing molecule-specific tag sequences (primer IDs), will be presented and discussed. The results will be presented in three main sections. 1) Performance of UDPS. 2) Dynamics of HIV-1 quasispecies. 3) Methods to reduce UDPS error frequency.

4.1 PERFORMANCE OF UDPS

4.1.1 Experimental setup

The UDPS technology has been used extensively during the last few years. For example to study cancer-associated genes in humans [245, 246] and minority variants within the HIV-1 quasispecies, including drug resistance (**Paper I**; [207, 238, 247-249], coreceptor use (**Paper III**; [240, 241, 244], APOBEC3 hypermutations [250] and coevolution in the *nef* gene [251]. Recently, whole genome deep-sequencing of HIV-1 has been described [252] as well as improvements in the UDPS accuracy by molecule-specific tags [253, 254]. The depth of UDPS depends on the number of viral templates that can be successfully extracted and amplified from the patient sample [207, 255], the error rate of PCR and UDPS, and the efficiency of cleaning the UDPS data from such errors. Different bioinformatic approaches have been developed to identify, remove or correct these sequencing artifacts, by in-house cleaning strategies (**Paper I**; **Paper IV**; [240-242]) or bioinformatic programs such as ShoRAH [243] and Segminator [244]. The approaches have been reported to decrease the average error rate to levels ranging from 0.05% (**Paper I**; [242]) to 0.43% [256] errors per nucleotide. Frequency estimates have been shown to be reproducible for variants constituting $\geq 1\%$ [241] and $>0.27\%$ (**Paper II**) of the population.

We have developed optimized pre-UDPS protocols (i.e. RNA extraction, cDNA and PCR) and we have quantified the number of HIV-1 cDNA templates subjected to UDPS, which ranged from 2,300 to 570,000 (**Paper I**) and 56 to 93,632 in **Paper III**. Such quantification verified that the number of templates was sufficient for deep sequencing. It should be noted that quantification of viral templates before UDPS has not always been used in other UDPS studies. If too few templates are recovered the depth of the UDPS analysis would be too low and a single genome sequencing approach [257] might be more useful. Furthermore, the quantification was done by a PCR system with the same sensitivity as the PCR used in the library preparation. Occasionally, we observed low efficiency of preparation, which might be explained by the fact that some samples had been suboptimally stored at -20°C for long time or repeatedly freeze-thawed. In **Paper I** and **Paper III**, 3,837 to 41,940 and 279 to 32,094 reads per sample were generated, based on template

molecule availability. Some samples were over-sampled, which means that the number of templates were lower than the number of reads. Over-sampling can theoretically remove some of the stochasticity in the distribution of variants. However, in cases of over-sampling it is important to remember that each read does not corresponds to an individual virus particle, which means that the sequence depth is affected. The maximal sequence depth of the samples with the lowest number of templates was 0.04% (1/2300) in **Paper I** and 1.8% (1/56) in **Paper III**. By analyzing the frequency and distribution of sequencing errors in experiments on plasmid clones we were able to develop bioinformatic scripts that were used to clean the data (**Paper I** and **Paper IV**) from sequencing artifacts and to determine statistical cut-off values for detection of high-confidence minority resistance mutations and genetic variants. The error rate across sites was estimated to be approximately 0.05% errors per nucleotide after data cleaning for both the *pol* amplicon (**Paper I** and **Paper IV**) and the V3 amplicon (**Paper III**) as well as for the 454 GS FLX and 454 GS FLX Junior Titanium platforms.

As expected, the error rate was not uniform across sites. For this reason we estimated the UDPS error rate for each drug resistance position (**Paper I**). These cut-off values were adjusted according to the number of reads generated for each sample. The average cut-off value for drug resistance mutations was estimated to be 0.05% (range 0.014 - 0.29%) and the average cut-off for high-confidence variants was estimated to be 0.11% (range 0.09 - 0.21%). We later observed (**Paper IV**) that the site-specific error rates in cleaned data were moderately, but still significantly, correlated between runs. There was also a moderate correlation between errors in forward and reverse reads of the same sample and run. Moreover, there were significantly more transition errors compared to transversion error after the data had been cleaned. Collectively the results from **Paper IV** indicate a many of the errors that remain after data cleaning were introduced during the PCR that preceded pyrosequencing. This means that it might be more correct to use individual cut-offs for transitions and transversions than the site-specific cutoffs that we used in **Paper I**.

In **Paper I**, **Paper II**, **Paper III** and **Paper IV** an average of 30%, 20%, 15% and 31% of the total number of reads were filtered during the data cleaning process, respectively. The lower percentage of reads removed in **Paper II** and **Paper III** is due to that no cut-off values were used. The average error frequency on raw reads was estimated to be 0.54% per nucleotide and the filtering strategy reduced the error rate approximately 10-fold. Thus, the filtering of 1/6 to 1/3 of the reads greatly improved the signal to noise ratio. The removal of reads with indels, which was mainly introduced during UDPS, had the greatest impact on reducing the error frequency. The sources of errors and the data filtering approach are discussed in more detail in section 4.3.1.

Taken together, after data cleaning the sensitivity of our UDPS methodology was primarily limited by errors introduced during PCR or the low number of templates for some samples.

4.1.2 UDPS evaluation

In **Paper II** the performance of UDPS was evaluated for experimental noise and data variability, such as repeatability, effects of sequence direction, sensitivity, influence of primer-related selective amplification and *in vitro* PCR recombination.

To evaluate the repeatability of frequency estimates of HIV-1 variants, we performed repeated UDPS analyses of two patient plasma samples (**Paper II**). We found that a repeated measurement had a 95% likelihood of lying within $\pm 0.5 \log_{10}$ of the initial estimate. Thus, a variant that was found in 100 reads in the first measurement had a 95% likelihood to lie between 32 and 316 reads in the second measurement. Interestingly, the repeatability was similar for rare and more abundant variants. We compared our results with those of Poon *et al.* [251], who used variance-to-mean ratios to investigate repeatability. The average variance-to-mean ratio in our experiments was 3.2×10^{-4} , which is more than 20 times lower than that estimated by Poon *et al.* [251]. In addition, Poon *et al.* reported that some variants representing 1 - 5% of the virus population in one analysis were not detected when the analysis was repeated. Similarly, Gianella *et al.* recently reported a low level of repeatability in detection and quantification of minority drug resistance mutations [258]. With our UDPS methodology, we repeatedly identified variants that represented $>0.27\%$ of the virus population. Thus, variants within the HIV-1 quasispecies that are as rare as 1 in 370 could be repeatedly detected. The reason for the differences in repeatability between these studies and our, is not clear, but could be due to differences in both laboratory methodology, sequencing approach and data cleaning. For instance Gianella *et al.* used a shotgun sequencing approach, which generally gives lower and less predictable sequence depth (coverage) than amplicon sequencing. One can also speculate that even lower frequencies could have been obtained if the samples would have been over-sampled, which was not the case in our study (**Paper II**).

In **Paper II**, we evaluated the effect of sequence direction on variant abundance estimates, which can be of importance in e.g. data cleaning. Bidirectional UDPS has been described in only a few studies (**Paper I**; [246, 256]), where variants were considered “true” if they were present in both sequence directions. We found that the difference in variant abundance between forward and reverse sequence direction in general was relatively small and approximately as great as the difference between UDPS runs (repeatability experiment described above). However, in contrast to the repeatability experiments, the agreement between forward and reverse analyses was higher for common variants than for rare variants, which was not surprising due to stochasticity in the ability to detect rare variants with abundance close to the detection limit. In fact, it was somewhat unexpected not to see this correlation in the repeatability experiment described above. In addition, some variants only exceeded our cut-offs for high-confidence variants in one sequence direction.

The sensitivity of our UDPS methodology to identify minority variants representing 0.5 and 0.05% of the population was evaluated in **Paper II** by using mixed known concentrations of two molecular clones. The minor variant was identified in both experiments, but their proportions were somewhat higher than expected, i.e. 2.2% and 0.31% respectively. This may be a stochastic effect, but we

cannot exclude the possibility that minority strains may have been systematically overestimated for instance if the major variants reached the PCR plateau earlier than rare variants. Artificial HIV-1 mixtures of 1% and 0.1% have previously been analyzed by Tsibris *et al.* [241] and Zagodi *et al.* [242], respectively. Our results are in agreement with these studies and suggest that it is possible to detect and quantify minor variants of the HIV-1 population, at least when the minor variant is clearly genetically distinguishable from the major variants such as is expected in the case of superinfection. If the variants are very similar (for instance a single transition) it is more challenging to differentiate between true variants and variants that have arisen due experimental error.

Moreover, in **Paper II**, the potential influence of primer-related selective amplification on estimation of variant abundance were evaluated using two nested primer sets with unique primer binding sites that targeted the same region in the *pol* gene. These primers targeted highly conserved, but separate, primer binding sites and included wobbled bases to further reduce the likelihood of nucleotide mismatches to the targets. Despite these efforts the estimations of variant abundance differed between the two primer sets. We were able to detect variants down to 0.2% of the viral population with both primer sets. However, one variant, which was estimated to represent 46% using the original primers, was detected in only 5.6% of the reads obtained with the alternative primers. As a result the limits of agreement was approximately two times wider than when the sample was re-analyzed with the original primer set. This suggests differential amplification of certain HIV-1 variants, presumably due to primer-related selective amplification. Thus, optimal primer design may be very important when UDPS is used to analyze the population structure in divergent target sequences, like HIV-1 populations. One could even speculate if multiple primer sets should be used in order to fully and correctly characterize HIV-1 variation.

UDPS has been used to study genetic variants and mutational linkage, but such analyses are only valid if the frequency of *in vitro* recombination is zero or close to zero. The most obvious source for recombination is the PCR, which is known be associated with incomplete extensions and template switching, which results *in vitro* recombination. The frequency of PCR recombination varies considerably with amplicon length and amplification condition [259-263]. To determine the *in vitro* recombination frequency in our experimental system we mixed two molecular clones in a 50:50 ratio before PCR amplification and UDPS. The two clones were selected so that they differed by 13 informative sites that were distributed across the amplicon. In addition, to study if the frequency of *in vitro* PCR recombination may be influenced by the number of target molecules we tested both 100,000 and 10,000 HIV-1 DNA templates as input in the outer PCR. The same dataset was used in **Paper I** and **Paper II**, but the analysis was extended in **Paper II**. Recombinant sequences were defined as sequences with replacement of at least two signature nucleotides that were adjacent (**Paper I**) or irrespective of whether they were adjacent or not (**Paper II**). The later definition is more conservative than the first and thus the estimated recombination frequencies were a little higher in **Paper II** compared with **Paper I**. In **Paper I** the estimated recombination frequency was 0.76% and 0.27% compared with 0.89% and 0.29% in **Paper II**, in the 100,000 and 10,000 template experiments, respectively. Importantly, the individual frequency of most of the recombination variants was

below the cut-off for high-confidence variants. Thus, these variants would not be considered as genuine in data analysis.

In summary, our results show that the repeatability of frequency estimates of HIV-1 variants was good for major as well as minor variants in patient plasma samples. This indicates that the experimental noise introduced during RNA extraction, cDNA synthesis, PCR and UDPS was low. However, for rare variants *in vitro* recombination and effects of sequence direction needs to be considered. Finally, the design of primers for PCR amplification is important during UDPS as well as for all PCR-based methods, since primer-related selective amplification can skew frequency estimates of genetic variants.

4.2 DYNAMICS OF HIV-1 QUASISPECIES

4.2.1 Minority variants

Minority HIV-1 variants have in recent reports shown to be of clinical significance, this applies especially to mutations associated with NNRTI resistance [206-212] and minority variants harbouring X4/R5X4 phenotype [217]. Such minority variants could either be transmitted or could have evolved over time within the patient. Theoretically, mutations associated with drug resistance are expected to occur naturally at low levels within the HIV-1 quasispecies, even if a patient has never received treatment [205]. By simple calculations, using the RT error rate of 1.4×10^{-5} mutations per site and generation [44] and viral production rate of 10^{10} virions per day [51] it can be estimated that all nucleotides in the HIV-1 genome mutate on average around 10^5 times per day in an HIV-1 infected individual. In **Paper I** and **Paper III**, we investigated the pre-existence of drug resistance in treatment naïve patients who later experienced treatment failure and pre-existence of X4-virus in patients whose virus later switches to CXCR4 use, respectively.

In **Paper I**, we found significant levels of the resistance mutations M184I (4 of 5 patients), T215I and/or T215A (4 of 5 patients) ranging from 0.02%–0.12% in plasma samples obtained before treatment was initiated. In contrast, we did not find any significant pre-existence of the drug resistance mutations M184V, Y181C, Y188C, L210 or T215Y/F. The presence of M184I, T215I and T215A in treatment naïve patients was somewhat expected since these drug resistance mutations only differ by one transition from wild-type (see discussion of **Paper IV**). For the same reason we would have expected to find M184V, Y181C and Y188C which also are transitions, but not T215Y/F since the latter are double mutants (containing transversions) compared to wild-type. One explanation for the absence of significant levels of M184V, Y181C and Y188C could be due to the higher cut-off values for high-confidence mutations at these positions (e.g. 0.15% for M184V compared to 0.07% for M184I). Since both the isoleucine (ATA) and valine (GTG) are due to transitions from the wild-type methionine (ATG), the higher cut-off seen in the M184V position is not explained by the transition vs. transversion error bias that we observed in **Paper IV**, but rather indicates presence of context-dependent errors (**Paper IV**). The T215I and T215A mutations involve two transversions compared to wild-type, thus their presence was somewhat unexpected. It should also be acknowledged that we cannot exclude that these patients were infected with resistant viruses which later reverted to the two T215 reversion mutations

T215I and T215A. The clinical relevance of the pre-existing drug resistance mutations found in our study is probably low. It has been shown that pre-existing M184I does not necessarily lead to virologic failure [212]. The same results have been shown for M184V mutation, which has been studied more extensively [206]. T215I and T215A do not by them self confer phenotypic resistance [186].

G->A mutations are the most common type of errors introduced by RT [44, 264], which appears to be the reason why M184I most often precedes the emergence of M184V during 3TC resistance *in vivo* (**Paper I**, [264]). This could explain why we found significant levels of pre-existing M184I in 4/5 patients (**Paper I**). In a previous study, the natural occurrence of M184V, Y181C, T215Y and T215F was estimated in samples obtained before the era of ARV to be 0.5%, 1.0%, 1.0% and 0.7%, respectively [265]. In **Paper I** we did not detect any preexistence of these mutations even though our method allowed detection of resistance mutations below the frequencies presented by Johnson *et al*. However, since we only studied five patients our study does not have power to draw conclusions about these differences.

In **Paper III**, we investigated if X4-viruses were present as a minority populations during PHI in three patients with HIV-1 populations that later switched coreceptor use to include CXCR4. UDPS analysis showed that the X4 variants that emerged in these three patients after coreceptor switch were not present during PHI and that the X4 variants most probably evolved from the R5 population during the course of infection rather than being transmitted as minor variants. However, we cannot exclude that X4 virus was transmitted and suppressed to levels below the detection limit of UDPS during the first weeks of infection. Thus, X4 virus could have been latent e.g. in the lymph nodes until the coreceptor switch. Nevertheless, if this was the case in these three patients we would probably have seen a greater genetic difference between the R5 and X4 variants that coexisted in the switch samples.

4.2.2 Virologic failure due to drug resistance

In **Paper I**, we investigated the development of drug resistance in six patients during treatment failure due to suboptimal treatment. Our UDPS analyses showed that different wild-type variants co-existed within each patient before initiation of therapy. Following start of therapy, virus variants with several different combinations of resistance mutations evolved and co-existed. However, during prolonged treatment failure the number of viral variants decreased, suggesting genetic bottle-necking due to a selective advantage of certain variants. This was accompanied by a gradual increase in the prevalence of variants with specific linked drug resistance mutations (in particular variants with M184V+T215Y and M184V+L210W+T215Y) and wild-type variants were only detected in one patient during therapy failure. However, wild-type variants might have been present below our detection limit of 0.11%, because it is well known that there is residual low level viremia (approximately 1 plasma HIV-1 RNA copy/ml) during long-term successful HAART. These viral variants often are drug-sensitive [161, 266] because they are released from stable reservoirs that were established before therapy was initiated [267]. It is reasonable to assume that such drug-sensitive variants are released also during non-suppressive HAART, but at such a low level that they were not detected in our UDPS analyses. In contrast to our results, Allers

et al. reported significant levels (0.6 to 30%) of lamivudine sensitive variants in viral population from patients failing 3TC-containing therapy [268]. However, these patients received dual-therapy with 3TC and AZT, while our patients generally received three or more drugs.

4.2.3 Treatment interruption and drug resistance

Since many primary drug resistance mutations reduce fitness [269], drug-sensitive viruses rapidly reemerge after complete treatment interruption [229, 270]. However, the kinetics and detailed dynamics of this process is incompletely understood. In **Paper I** we showed that drug resistant variants decreased to undetectable levels a few months after complete treatment interruption even when highly sensitive UDPS was used. This indicates that these variants had quite low fitness in absence of therapy. The complete out-growth of drug-sensitive variants within a few months differs markedly from the findings in patients with transmitted drug resistance, where drug resistance may persist for many years [226, 271, 272]. Our findings might be of clinical relevance since we show that drug resistant variants may be very difficult to detect in patients with previous treatment failure even with highly sensitive UDPS technology. In contrast to our results, Le *et al.* found low abundance mutations associated with AZT/d4T resistance 2 to 7 years after treatment with these drugs had been stopped [247]. However, in contrast to our patients, the patients enrolled in their study continued therapy with other antiviral drugs. Additional studies are needed to investigate the dynamics of drug resistant minority variants after treatment interruption. Since drug resistant variants can become deposited in long term reservoirs [271, 273] it would be interesting to analyze different cell compartments in addition to plasma.

4.2.4 Coreceptor switch

X4 or X4R5 viruses emerge in about 70% of infected individuals in the later stages of infection [77, 110-112] by stepwise mutational pathway from R5 variants [274, 275]. In **Paper III**, the coreceptor switch was studied in three patients. Our results show that the coreceptor switch was relatively abrupt rather than gradual since the preceding samples obtained 15-20 months prior to the switch contained only R5 virus. Phylogenetic analyses showed that distinct R5 and X4 clusters were present after the switch in all patients. In two of the patients X4 variants appeared to have emerged from the contemporary R5 population generating two or more statistically supported X4 clusters, which also supports the notion that the R5 to X4 switch is relatively quick as indicated by Bunnik *et al.* [274].

4.2.5 Primary HIV infection

Several recent studies have indicated that most HIV-1 infections are established by one or a small number of infectious particles [70, 72, 73], resulting in low genetic diversity during PHI and a gradually increase during the asymptomatic course of infection [75, 78, 276, 277]. However, it is difficult to distinguish if the infection actually is established by several founder viruses, but only one or a few are selected due to replicative capacity, ability to bind to and infect host cells, number of virions produced or recognition by the immune system. Similar to the studies above, we observed one to three major variants during PHI of three patients (**Paper III**). The viral diversity during PHI in our study was 1.1 - 3.2%,

which is within the upper range of what has been reported before where the median viral diversity was 0.39% (range 0.04% - 3.23%) [276]. However, it is difficult to make exact comparisons between genetic diversities reported in these studies since different sequence methodologies have been used and the exact timing of sampling was different. Deep sequencing will naturally have the possibility to give a higher average genetic distance due to the higher coverage of variants compared with methods based on conventional Sanger sequencing techniques such as single genome sequencing.

4.3 METHODS TO REDUCE UDPS ERROR FREQUENCY

4.3.1 UDPS errors and data cleaning approach

In **Paper IV**, we investigated the frequency and type of errors that occur during UDPS of a molecular clone corresponding to the same fragment of the HIV-1 *pol* gene analyzed in **Paper I** and **Paper II**. Based on these results we developed a strategy to clean UDPS data from sequencing errors. The cleaning strategy reduced the average error frequency per base from 0.54% to 0.055%, i.e. by a factor of 9.8. UDPS errors are known to be over-represented in homopolymeric regions [246, 256, 278], but this bias was removed by our data cleaning strategy. However, the UDPS errors that remained after data cleaning were 59-times more likely to be transitions than transversions, which indicates that a substantial proportion of the errors were introduced during PCR. This conclusion was corroborated by the finding of moderate, but still highly significant, correlation of sequencing errors observed in forward and reverse sequencing direction of the same amplicon. If the errors were introduced during UDPS, the probability to detect the same error in both forward and reverse direction would have been lower. The transition vs. transversion bias needs to be considered when cut-offs for detection of minority mutations are determined. Thus, it would have been more correct to base the cut-offs used in **Paper I** on transitions and transversions instead.

In agreement with previous reports, we found that errors in raw UDPS data were over-represented in homopolymeric regions (**Paper IV**). Indels, especially deletions, was the most frequent error type [279], but substitution errors also occurred. Indels are mainly created *in silico* during the interpretation of the intensity of chemiluminescent signal in homopolymeric regions [280]. In contrast, substitution errors may be introduced both in the PCR and in the UDPS [242, 246]. Our cleaning strategy did not manipulate the sequences and thereby did not create artificial new variants. However, it cannot be excluded that the cleaning strategy may occasionally have filtered genuine sequence variants when it was applied on UDPS data from patient samples. We observed that the error frequencies per nucleotide were unevenly distributed across the sequence fragment. To understand in which steps of the sequencing procedure the errors had been introduced, we investigated the characteristics of the substitution errors that remained after data cleaning. We found that the site-specific error frequencies in forward and reverse sequencing direction of each of the three runs were significantly correlated, which indicates that a proportion of the errors were introduced in the PCR. As presented in the results section (**Paper IV**), such correlation is only expected for errors that are introduced in early PCR cycles, which means that the proportion of PCR errors that remained after data

cleaning may have been higher than indicated by the correlation coefficients, which were relatively low. There was also a significant correlation of site-specific error frequencies between runs, which indicates presence of context-dependent errors that may have occurred in the PCR and/or UDPS.

4.3.2 Reducing errors by molecule-specific tags

In **Paper IV** we describe that the depth of UDPS is limited by errors introduced during PCR and/or UDPS. Here, preliminary results from “The molecule-specific tag project” are described. In this project we have attempted to improve the UDPS methodology to allow more accurate detection of minority viral variants. We have used this method to amplify the same fragment of 167 nucleotides in the *pol* gene as in **Paper I**, **Paper II** and **Paper IV**, however the method is generic and other genes/organisms can be targeted. The method is based on the following key features: a reverse primer with a unique sequence tag of 10 degenerated nucleotides ($4^{10} = 1,048,576$ unique combinations) is added in the cDNA synthesis, giving each individual template a specific genetic barcode (tag). Next, a one cycle PCR generates double stranded DNA using platinum taq high fidelity enzyme. The uracil containing reverse primer is degraded by uracil-DNA glycosylase and NaOH. The double stranded DNA template is amplified with a semi-nested PCR and subsequently pooled before UDPS. The generated UDPS reads are sorted based on their sample and molecule tags. At least three reads is needed from each template to create a consensus sequence, which results in almost complete elimination of experimental errors that have occurred after cDNA synthesis, i.e. during PCR and UDPS. After we had initiated our project other groups have published methods that apply the same basic principle, i.e. tagging of template molecules to allow them to be re-sequenced [253, 254]. Jabara *et al* referred to the template-specific genetic barcodes as primer IDs.

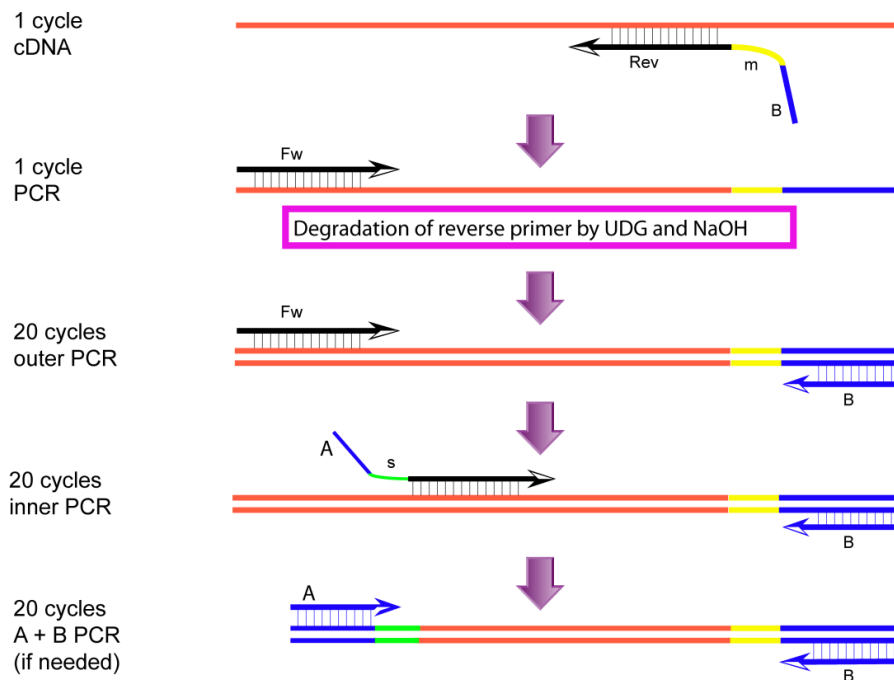


Figure 13. Experimental setup. Rev, reverse primer. m, molecule-specific tag (primer ID). B, B adaptor. UDG, Uracil-DNA glycosylase. Fw, forward primer. A, A adaptor. s, sample-specific tag.

Five patients with documented TDR (based on Sanger sequencing) and one plasmid clone sample were selected for sequence analysis using the primer ID approach. A total of 408,717 UDPS reads, of which 41,387 reads originated from the plasmid clone were generated. The average error frequency per nucleotide for the plasmid sequences without consideration of primer IDs was estimated to be 0.6% (which is similar to the raw error rate in previous UDPS runs, **Paper IV**). The reads originating from each primer ID were aligned and a consensus sequence was generated for each template. A total number of 162 consensus sequences were generated for the clone. An example of an alignment from five reads with the same primer ID is shown in Figure 14. The alignment contains both substitution errors and deletions but since these errors are in minority (<50% of the nucleotides each position) they are not present in the consensus sequence, which represents the “error-free” sequence of the template molecule. After creating a consensus sequence for each unique primer ID sequences the error frequency was reduced five-fold to 0.13%, without any prior cleaning of the reads.

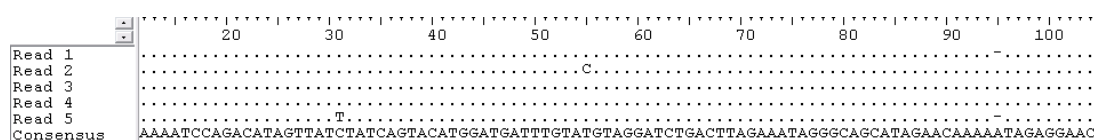


Figure 14. An alignment of reads originating from the same template with a unique molecule-specific tag sequence. The variation seen in the alignment are due to errors introduced during PCR or sequencing. The generated consensus sequence contains no errors.

There are several reasons why the error frequency of the method was not 0%. As previously shown, deletions were the most common error, followed by insertions and substitutions (**Paper IV**). Most of the deletions and insertions were localized in homopolymeric regions. For a small number of template molecules >50% of the reads had an identical error in a homopolymeric region (usually a deletion), which generated an incorrect consensus sequence of this template molecule. By correcting these indels in homopolymeric regions the error frequency was reduced to 0.06% errors per nucleotide.

Eight unique consensus sequences still contained errors relative to the plasmid sequence. These errors were substitution errors (n=4) or insertions (n=3) in non-homopolymeric regions that could have three sources: 1) The same type of systematic UDPS error as described above; 2) A PCR error in the first PCR cycle; or 3) Polymorphisms in the plasmid templates. We observed a high substitution rate at one position where over 10% of the reads harbored a T->C transition (Figure 15, lower graph). When studying the individual alignments the mutation was present in all reads originated from nine template molecules and was uncommonly found in other templates. Thus, it is unlikely that this mutation originated from PCR or UDPS errors due to the low probability of generating the same substitution error at the same site at nine separate templates. In addition, we have not observed similar result in any previous studies using this *pol* fragment (**Paper I**, **Paper II** and **Paper IV**). Hence, a more likely explanation for this error was a polymorphism in plasmid. When these sequences were removed the error frequency was reduced to 0.0006% errors per nucleotide.

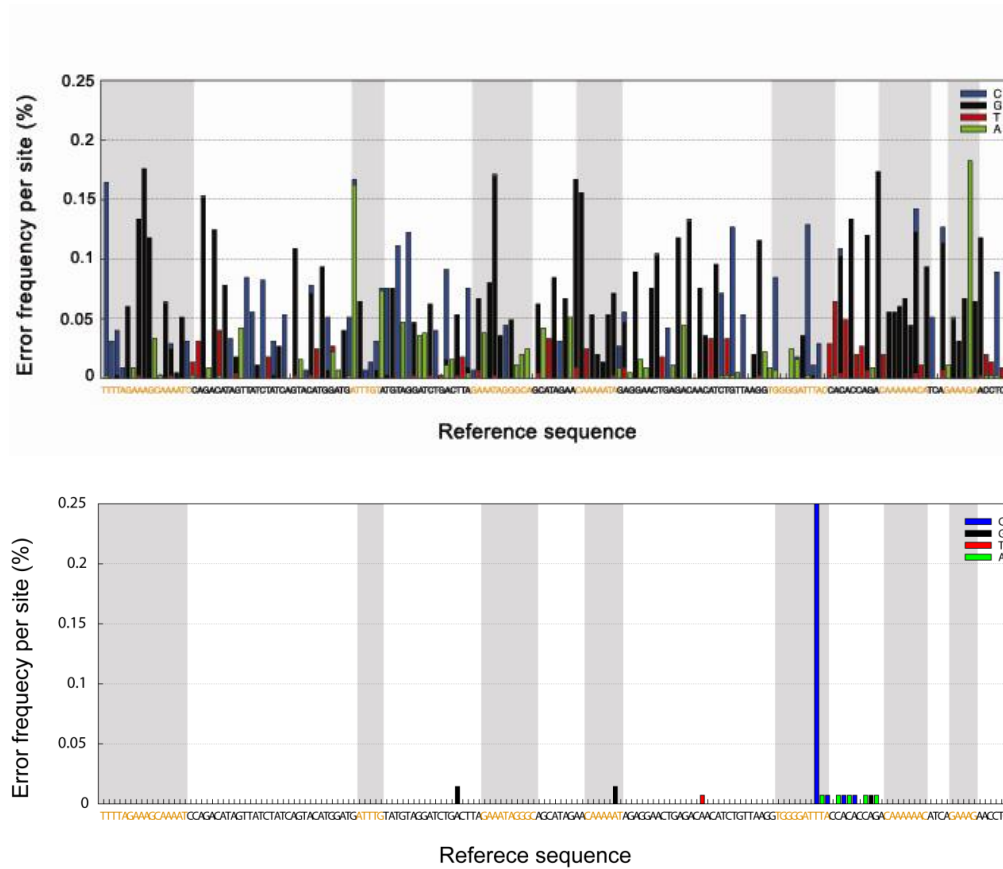


Figure 15. Error frequency per nucleotide. The upper graph shows the error rate per site for the UDPS methodology used in Paper I, Paper II, Paper IV, after data cleaning. The lower graph shows the error rate per site using molecule specific tags after consensus formation. The bars are color-coded according to the substitution error. Homopolymeric regions are shaded.

It needs to be emphasized that the 7 unique consensus sequences containing errors were still present and that they constituted 4% of the consensus variants (7/162), shown in Figure 16. This was higher than expected to be generated from first two single PCR cycles, which theoretically would generate an error in 0.14% of templates (4.33×10^{-6} errors/bp \times 167bp \times 2 PCR cycles). In addition, the fact that 3/5 of the substitution errors were transversions, which are uncommon as PCR errors (**Paper IV**), further indicates that the origin of these errors are more complex and require further investigation.

We are currently analyzing the UDPS reads from the patient samples, generating consensus sequences for each template, to accurately quantify minor TDRs, such as M184V. Our preliminary results show that this method has the potential to increase the depth of UDPS by reduced error frequency per nucleotide from 0.6% down to 0.0006%. Furthermore, the method gives an exact count of the number of analyzed templates which eliminates the need for independent template quantification as well as problems with unintentional re-sampling of the same templates. However, we and Jabara *et al.* have observed an unbalance in the number of reads per template [253], which might be due to that some primer IDs are more favorable for amplification than others. Moreover, in our data we observed PCR errors in primer IDs, which can lead to false templates. However, by grouping the primer IDs based on similarity, thus merging primer IDs that only differ by one or two nucleotides from a common primer ID, the risk of generating false templates can be reduced. In addition, we have observed

lower number of primer IDs than expected in the clone data. From the 10,000 plasmid templates subjected to the experiment, only 162 unique templates were detected. This was due to three main reasons: 1) A lower recovery of templates throughout the experimental setup than expected. 2) Skewness of amplification of individual primer IDs due to secondary structure or other factors making some primer IDs more favorable for amplification. 3) A lower total number of reads generated from the 454 GS FLX titanium run than expected. Thus, several primer IDs were found in fewer than three reads and were discarded since no consensus could be constructed.

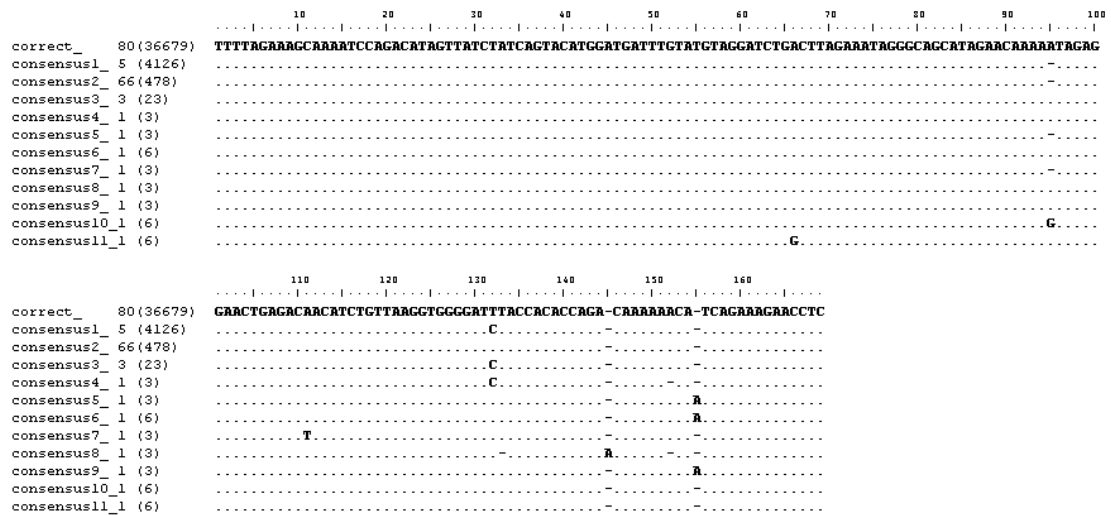


Figure 16. Alignment of the 12 unique consensus sequences representing individual plasmid template. The first consensus sequence shows the correct plasmid sequence in which was observed in 80 consensus sequences. The number of reads is shown in brackets. Consensus sequence 1, 3 and 4 harbored the T->C constituting more than 10% of the reads.

Taken together, we and others [253, 254] have shown that the addition of primer IDs to tag each template is a possible solution to remove artificial errors introduced during PCR, which is a step needed for all UDPS method used today. This method has the potential to increase the sequence depth and allow more accurate studies of the HIV-1 quasispecies.

5 CONCLUSIONS AND FUTURE PERSPECTIVES

HIV-1 has the ability to quickly diversify and adapt to changes in its' environment, such as evading the immune response of the host [12], altering cell tropism, and developing resistance to antiretroviral drugs [13]. In this thesis the UDPS technology has been used to dissect the HIV-1 quasispecies of HIV-1 infected patients to study development of drug resistance and evolution of cell tropism. The UDPS methodology has been carefully optimized to maximize the depth and accuracy of our analyses.

We and others have used the UDPS technology to study minority variants within the HIV-1 quasispecies, in regards to drug resistance (**Paper I**; [207, 238, 247-249], coreceptor use (**Paper III**; [240, 241, 244], APOBEC3 hypermutations [250] and coevolution in the *nef* gene [251]. The depth of UDPS depends on the number of viral templates that can be successfully extracted and amplified from a plasma sample [207, 255], the error rate of PCR and UDPS, and the efficiency of cleaning the UDPS data from such errors. Thus, an experimental design that allows high recovery of HIV-1 templates together with an effective data cleaning strategy is important for successful UDPS analyses (**Paper I, Paper IV**). Different bioinformatic cleaning approaches have been reported to decrease the average error rate to levels ranging from 0.05% (**Paper I, Paper IV**; [242] to 0.43% [256] errors per nucleotide. Variant abundance estimates has been shown to be reproducible for variants constituting $\geq 1\%$ [241] and $>0.27\%$ (**Paper II**) of the population.

In **Paper II**, we have performed a series of experiments to evaluate the performance of our UDPS analysis. The results showed that the repeatability was good for major as well as minor variants in patient plasma samples, which indicates that the experimental noise introduced during RNA extraction, cDNA synthesis, PCR and UDPS was low. However, for rare variants *in vitro* recombination and effects of sequence direction needs to be considered. Finally, the design of primers for PCR amplification is of special importance during UDPS, since primer-related selective amplification can skew frequency estimates of genetic variants. However, it remains to be investigated if our results can be generalized to other gene fragments or longer read length.

In **Paper I**, we showed that the levels of pre-existing drug resistance in plasma samples from five treatment naive patients was very low and that several important drug resistance mutations (M184V, Y181C, Y188C and T215Y/F) were not detectable in pre-treatment samples, indicating that the natural occurrence of these mutations were below our detection limit. However, we found low, but significant, levels of M184I (4 of 5 patients), T215I and/or T215A (4 of 5 patients) at proportions ranging from 0.02%–0.12%. The clinical significance of these mutations is probably low. It has been shown that pre-existing M184I does not necessarily lead to virologic failure [212] and that the T215I/A mutations do not by them self confer phenotypic resistance [186].

During treatment failure and treatment interruption, we found almost 100% replacement of wild-type and drug-resistant variants, respectively (**Paper I**). This

implies that the proportion of minority variants with drug resistance in patients with previous treatment failure can be too low to be detectable even with highly sensitive UDPS technology. For optimal treatment management of such patients it would be interesting to investigate the utility of analyzing viral DNA in PBMCs.

In **Paper III**, three patients with HIV-1 populations that switched coreceptor use were investigated. UDPS analysis showed that X4 virus that emerged after coreceptor switch was not detected during PHI and that the X4 population most probably evolved from the R5 population during the course of infection rather than was transmitted as minor variants. Moreover, one to three major variants were found during PHI, lending support to the hypothesis that infection usually is established with one or just a few viral particles [70, 72, 73].

We have investigated the frequency and type of errors that occurred during UDPS (**Paper IV**). The errors that remained after data cleaning were significantly more often transitions than transversions, which indicates that a substantial proportion of these errors were introduced during PCR. This affects the limits of detection of minority mutations since UDPS analyses of HIV-1 are presided by a PCR step.

To circumvent these errors an improved methodology was developed with the intention to allow more accurate detection of minority viral variants. In this method each HIV-1 template was given a specific genetic barcode (primer ID) prior to the PCR and by subsequently generating at least three sequences from each template, consensus sequences with minimal errors can be constructed. Recently similar approaches have been described [253, 254]. Our preliminary results showed a reduced UDPS error frequency from 0.6% in raw reads to 0.0006% errors per nucleotide after consensus generation. This improved methodology has the potential to increase the sensitivity and accuracy of UDPS analyses 1000-fold. Taken together, our studies show that UDPS can be used to gain new insights in HIV evolution and resistance and is relevant for the possible future clinical use of this technology.

Current routine HIV-1 resistance testing is performed by population Sanger sequencing, which has the disadvantage of only detecting mutations present in >20% of the virus population. However, it has been shown that minority variants below this detection limit may have clinical relevance. This applies especially to minority NNRTI mutations [206-212]. Specifically, the presence of minority variants representing >0.5% of the viral population conferred a significant higher risk of virologic failure compared with minority variants present at less than 0.5% [206]. Because NNRTI-based regimens are the most commonly prescribed first-line therapy, the clinical use of a more sensitive method could help identify individuals at increased risk of virologic failure and thereby improve clinical management. One solution could be to use a real-time PCR method to detect the presence of minority drug resistance mutations. However, due to the high number of drug resistance mutations needed to be tested a whole-genome NGS sequencing approach might be more cost-effective. Recently, whole genome deep sequencing of HIV-1 has been described [252]. This approach could have great potential for improving the sensitivity of resistance tests used in the clinic and it is likely that Sanger sequencing for HIV-1 drug resistance will be outcompeted relatively soon. However, for other application such a single genome sequencing (SGS) [257] of samples with low copy number, Sanger sequencing will only be outcompeted

when technologies offering longer and more accurate reads have become available.

For routine drug resistance testing linkage between mutations is not crucial, thus the HIV-1 genome could be randomly fragmented and sequenced on any of the NGS platforms. To increase the accuracy of the NGS method it might be advisable to combine this method with a primer ID approach to tag individual templates. To be able to select which NGS technology would be most suitable several aspects need to be considered, such as sequence depth, evenness of coverage, read length, read quality, running costs, simplicity of workflows, total run times and scalability. A recent study by Loman *et al*, compared the three high throughput benchtop instruments available today (454 GS junior/Roche, MiSeq/Illumina, Ion Torrent/Life Technologies). The MiSeq and Ion Torrent had the highest throughput. The 454 GS Junior generated the longest reads and most contiguous assemblies but had the lowest throughput. MiSeq had the lowest error rates and the Ion Torrent and 454 GS Junior both produced homopolymer-associated indel errors. The number of indel errors was higher for Ion Torrent compared with 454 GS Junior. Moreover, the Ion Torrent had the shortest run time [279]. This study demonstrates that each technology has a trade-off between advantages and disadvantages.

A NGS approach could be used for routine type/subtype determination of other viruses (influenza and HCV) or for investigation outbreaks of bacterial pathogens. However, for type/subtype determination the depth does not need to be as great as for drug resistance testing instead it is important to generate as good coverage of the genome as possible to be able to make a correct assembly of the whole genome. This could also be used to determine recombinant virus variants and new strains.

The length of the sequence reads becomes crucial for deep sequencing projects where individual HIV-1 variants or the linkage between mutations are studied. Examples of such studies are: understanding the mechanism behind drug resistance development and coreceptor switch or identifying and characterizing distinct viral sub-populations in different compartments within HIV-1 patients. These kinds of projects are usually more research-based and do not have an obvious connection to routine tests in the clinic. The NGS method which allows the longest read length is 454 sequencing which allows ~400-500 bases. However, the absolute length might vary between different amplicons due to differences in nucleotide structure. By sequencing as long reads as possible the risk of generating artificial variants are reduced compared with shotgun approaches using short reads and assembly algorithms. The accuracy of the 454 assay will be improved by the use of a primer ID approach (as described above) to create consensus sequences for each template. The potential for the use of the primer IDs in deep sequencing studies is high and when the read length of the 454 sequencing technology increases, this method more and more mimics a SGS method but with a high throughput. Nonetheless, to optimally use the primer ID approach further investigation is needed to overcome skewness of template amplification and optimization of experimental setup to increase recovery. Furthermore, methods that allow direct sequencing of cDNA or ideally RNA without the need of prior PCR amplification would be the optimal choice for deep sequencing projects of HIV-1 especially if errors introduced during sequencing

and cDNA synthesis could be removed by e.g. a triplicate sequence of each template (similar to the primer ID approach). Today we are only in the beginning of exploring the future potential for NGS both in the clinic and in research based settings, exiting next coming years are to be expected. Thus, it is very likely technological advances will continue to allow better and better insights into the evolution of HIV-1 and other pathogens.

6 ACKNOWLEDGEMENTS

There are many people that I would like to thank who have contributed to the work in this thesis and who have supported in many ways during my time as a PhD Student.

First I would like to say thank you to all patients participating in these studies. Without their contribution, this work would not have been possible.

Jan Albert: I am amazed by your knowledge and expertise and I am grateful for having been a part of your research team. Thank you for always being available for questions and for taking time to discuss my work. Thank you for always being positive and looking at research results from the bright side. I am immensely grateful for all the support you have shown me during my years as a PhD student.

Mattias Mild: Thank you for becoming my co-supervisor and friend. Thank you for your enthusiasm, passions and for sharing your knowledge in HIV and phylogenetics. Thank you for all support, great discussions, energy and laughs throughout the years. Thank you for encouraging me in times when I needed it!

Joakim Lundeberg: for introducing me to 454 sequencing and for scientific discussions.

Annika Karlsson: for input and scientific discussions and for being my co-supervisor.

Thomas Leitner: for great scientific discussions. Your knowledge and input has been very valuable.

Sarah Palmer: You have a true research soul and your enthusiasm and energy for science is an inspiration. Thank you for sharing your scientific knowledge and for all talks, dinners and fun times. Thank you so much for introducing me to interesting people.

Sven Britton: For showing me Ethiopia, one of the most interesting trips I have ever done. Thank you for your friendship and interesting discussions.

Göran Bratt: For introducing me to the clinical work at Venhälsan and letting me meet patients.

Richard Neher: for interesting collaborations and valuable scientific discussions.

Mats Nilsson and Lotte Moens: For interesting discussion and fruitful collaborations.

Emilie Hultin, Sara Arroyo Mühr and Carina Eklund: for exceptional guidance and support during the 454 GS Junior experiments.

Tomas Johansson in Lund: Thank you for all your advices regarding the 454 titanium experiment.

Joakim Esbjörnsson: for sharing your nice illustrations!

To the Jan Albert's group at SMI. **Afsaneh Heidarian**, for taking care of the lab and for being such a nice person and friend. **Kajsa Apéria**, for introducing me to P3 lab and all your help in the lab. **Maria Axelsson**, for all support and talks. **Marianne Jansson**, for all support. **Lisbeth Löfstrand**, your help with the administration has been very valuable.

Benita Zweyberg Wirgart: For the nice welcoming atmosphere at the clinical microbiology department.

Eva Ericsson: I am so grateful that I have got the chance to work with you and I am impressed by your organization and lab skills.

My wonderful colleagues and friends. **Johanna Brodin**, for bringing your bioinformatic expertise into the group. Thank you for being a great colleague and for becoming a true friend. I hope we will work together again in the future. **Wendy Murillo**, my Honduran friend. Thank you for a true friendship, I have really enjoyed all our lunches, dinners and talks. I hope to come and visit you soon in Honduras. **Melissa Norström**, for your friendship, all laughs and for all crazy things we have done together. I hope to see more of you soon. **Salma Nowroozalizadeh**, for your friendship and for all fun things we have done together. I hope to see more of you now when you are back in Sthlm! **Helena Skar**, for all your support and help throughout the years, thank you! **Carina Perez**, for being such an inspiring person. **Leda Parham**, for your friendship. **Lina Josefsson**, for your friendship and all interesting discussion about the future. Thank you for founding The PhD club and for letting me be a part of the team. **Susanne Eriksson**, for crazy shopping time in Seattle, I hope we will do more of that in the autumn. **Linda Trönnberg**, for friendship, lunches and dinners. **Alex Heddini**, for being my student and for your enthusiasm about the project despite the "disaster". **Tara Wahab**, for all nice discussions and for a great time in Ethiopia. **Malin Stoltz**, for your friendship. To all the other students: **Dace Balode**, **Viktor Dahl** and **Marcus Buggert**: It has been a pleasure to meet you and to have lunches and scientific discussions together.

To the other members of The PhD club: Cofounder **Therese Högfeldt**, for your friendship and all your support and long talks. **Cecilia Jädert**, for great collaboration! I will miss The PhD Club and I hope that I always can be part of it somehow ☺

Till alla mina fantastiska och underbara **vänner** utanför den akademiska världen. Jag kan inte nämna alla vid namn för det skulle ha fyllt många sidor men vet att ni alla betyder så oerhört mycket för mig, jag har mycket att ta igen med er! Speciellt tack till **Emma** och **Joakim** för er vänskap och för att ni med glädje har tagit hand om Vilhelm då tiden inte har räckt till. **Linda Bengtsson** för sann vänskap trots långt geografiskt avstånd. **Elle**, **Linda S**, **Maarit**, **Anna F**, **Camilla** och **Sara**, jag vill ses mer!

Mona och **Göran Gustafsson**: Tack för allt ert stöd och för att ni är de bästa farföräldrarna i världen för Vilhelm. Jag är så oerhört tacksam för allt som ni har gjort och för att jag har fått lära känna er. Utan allt ert stöd hade denna avhandling inte blivit klar på många år.

Mamma och **pappa**: Tack för att ni alltid har trott på mig och låtit mig gå min egen väg. Jag älskar er!

Louise: Tänk att vi valde samma spår. Tack för att du kan läsa min tankar och förstår mig som bara en tvillingsyster kan göra. Du är min bästa vän!

Petter: Tack för all kärlek! Tack för att du så osjälviskt fixar allt och för att du har tagit alla nätter! Tack för att du är världens bästa pappa! Tack för att du har läst min avhandling och din hjälp med figurerna! Jag ser fram emot nya spännande äventyr tillsammans, men först har jag så klart mycket att ta igen ☺

Vilhelm: Min älskade lille kille, den kärlek jag får från dig är obeskrivlig och gör allt känns möjligt.

7 REFERENCES

1. Gao, F., et al., *Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes*. Nature, 1999. **397**(6718): p. 436-41.
2. Gao, F., et al., *Human infection by genetically diverse SIVSM-related HIV-2 in west Africa*. Nature, 1992. **358**(6386): p. 495-9.
3. Keele, B.F., et al., *Chimpanzee reservoirs of pandemic and nonpandemic HIV-1*. Science, 2006. **313**(5786): p. 523-6.
4. Worobey, M., et al., *Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960*. Nature, 2008. **455**(7213): p. 661-4.
5. Korber, B., et al., *Timing the ancestor of the HIV-1 pandemic strains*. Science, 2000. **288**(5472): p. 1789-96.
6. CDC, *Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men – New York city and California*. MMWR Morb Mortal Wkly Rep 1981. **30** p. 305-308.
7. Hymes, K.B., et al., *Kaposi's sarcoma in homosexual men-a report of eight cases*. Lancet, 1981. **2**(8247): p. 598-600.
8. CDC, *Update on acquired immune deficiency syndrome (AIDS)--United States*. MMWR Morb Mortal Wkly Rep, 1982. **31**: p. 507-8, 513-4.
9. CDC, *Possible transfusion-associated AIDS*. MMWR Morb Mortal Wkly Rep 1982. **31**: p. 652-654.
10. Barre-Sinoussi, F., et al., *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)*. Science, 1983. **220**(4599): p. 868-71.
11. Holmes, E.C., *What does virus evolution tell us about virus origins?* J Virol, 2011. **85**(11): p. 5247-51.
12. Nowak, M.A., et al., *Antigenic diversity thresholds and the development of AIDS*. Science, 1991. **254**(5034): p. 963-9.
13. Kantor, R., et al., *Evolution of resistance to drugs in HIV-1-infected patients failing antiretroviral therapy*. AIDS, 2004. **18**(11): p. 1503-11.
14. Takehisa, J., et al., *Origin and biology of simian immunodeficiency virus in wild-living western gorillas*. J Virol, 2009. **83**(4): p. 1635-48.
15. Lemey, P., et al., *Tracing the origin and history of the HIV-2 epidemic*. Proc Natl Acad Sci U S A, 2003. **100**(11): p. 6588-92.
16. Worobey, M., et al., *Island biogeography reveals the deep history of SIV*. Science, 2010. **329**(5998): p. 1487.
17. de Sousa, J.D., et al., *High GUD incidence in the early 20 century created a particularly permissive time window for the origin and initial spread of epidemic HIV strains*. PLoS One, 2010. **5**(4): p. e9936.
18. UNAIDS, *Report on the global aids epidemic 2010*. Geneva, 2010.
19. UNAIDS, *AIDS Epidemic Update 2009*. Geneva, 2009.
20. UNAIDS, *World AIDS day report 2011*. Geneva, 2011.
21. Skar, H., C. Hedskog, and J. Albert, *HIV-1 evolution in relation to molecular epidemiology and antiretroviral resistance*. Ann N Y Acad Sci, 2011. **1230**: p. 108-18.
22. Freed, E.O., *HIV-1 gag proteins: diverse functions in the virus life cycle*. Virology, 1998. **251**(1): p. 1-15.
23. Dalgleish, A.G., et al., *The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus*. Nature, 1984. **312**(5996): p. 763-7.
24. Freed, E.O., *HIV-1 and the host cell: an intimate association*. Trends Microbiol, 2004. **12**(4): p. 170-7.
25. Berger, E.A., P.M. Murphy, and J.M. Farber, *Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease*. Annu Rev Immunol, 1999. **17**: p. 657-700.
26. Hu, W.S. and H.M. Temin, *Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination*. Proc Natl Acad Sci U S A, 1990. **87**(4): p. 1556-60.
27. Balakrishnan, M., et al., *Template dimerization promotes an acceptor invasion-induced transfer mechanism during human immunodeficiency virus type 1 minus-strand synthesis*. J Virol, 2003. **77**(8): p. 4710-21.
28. Schroder, A.R., et al., *HIV-1 integration in the human genome favors active genes and local hotspots*. Cell, 2002. **110**(4): p. 521-9.
29. Vatakis, D.N., et al., *Immediate activation fails to rescue efficient human immunodeficiency virus replication in quiescent CD4+ T cells*. J Virol, 2007. **81**(7): p. 3574-82.
30. Plesa, G., et al., *Addition of deoxynucleosides enhances human immunodeficiency virus type 1 integration and 2LTR formation in resting CD4+ T cells*. J Virol, 2007. **81**(24): p. 13938-42.

31. Agosto, L.M., et al., *HIV-1 integrates into resting CD4+ T cells even at low inoculums as demonstrated with an improved assay for HIV-1 integration*. Virology, 2007. **368**(1): p. 60-72.
32. Loisel-Meyer, S., et al., *Glut1-mediated glucose transport regulates HIV infection*. Proc Natl Acad Sci U S A, 2012. **109**(7): p. 2549-54.
33. Rambaut, A., et al., *The causes and consequences of HIV evolution*. Nat Rev Genet, 2004. **5**(1): p. 52-61.
34. Sodroski, J., et al., *Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat*. Science, 1985. **227**(4683): p. 171-3.
35. Strebel, K., *Virus-host interactions: role of HIV proteins Vif, Tat, and Rev*. AIDS, 2003. **17 Suppl 4**: p. S25-34.
36. Virology, F., *Fifth edition*. 2007. **2**: p. 2165.
37. Mariani, R., et al., *Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif*. Cell, 2003. **114**(1): p. 21-31.
38. Chiu, Y.L. and W.C. Greene, *The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements*. Annu Rev Immunol, 2008. **26**: p. 317-53.
39. Newman, E.N., et al., *Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity*. Curr Biol, 2005. **15**(2): p. 166-70.
40. Jiang, J. and C. Aiken, *Maturation-dependent human immunodeficiency virus type 1 particle fusion requires a carboxyl-terminal region of the gp41 cytoplasmic tail*. J Virol, 2007. **81**(18): p. 9999-10008.
41. Peng, C., et al., *Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity*. J Virol, 1989. **63**(6): p. 2550-6.
42. Vogt, V.M., *Proteolytic processing and particle maturation*. Curr Top Microbiol Immunol, 1996. **214**: p. 95-131.
43. Mansky, L.M. and H.M. Temin, *Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase*. J Virol, 1995. **69**(8): p. 5087-94.
44. Abram, M.E., et al., *Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication*. J Virol, 2010. **84**(19): p. 9864-78.
45. Drake, J.W., et al., *Rates of spontaneous mutation*. Genetics, 1998. **148**(4): p. 1667-86.
46. Zhuang, J., et al., *Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots*. J Virol, 2002. **76**(22): p. 11273-82.
47. Jetzt, A.E., et al., *High rate of recombination throughout the human immunodeficiency virus type 1 genome*. J Virol, 2000. **74**(3): p. 1234-40.
48. Levy, D.N., et al., *Dynamics of HIV-1 recombination in its natural target cells*. Proc Natl Acad Sci U S A, 2004. **101**(12): p. 4204-9.
49. Neher, R.A. and T. Leitner, *Recombination rate and selection strength in HIV intra-patient evolution*. PLoS Comput Biol, 2010. **6**(1): p. e1000660.
50. Ho, D.D., et al., *Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection*. Nature, 1995. **373**(6510): p. 123-6.
51. Perelson, A.S., et al., *HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time*. Science, 1996. **271**(5255): p. 1582-6.
52. Ho, D.D., *Perspectives series: host/pathogen interactions. Dynamics of HIV-1 replication in vivo*. J Clin Invest, 1997. **99**(11): p. 2565-7.
53. Allen, T.M., et al., *Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution*. J Virol, 2005. **79**(21): p. 13239-49.
54. Frost, S.D., et al., *Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection*. Proc Natl Acad Sci U S A, 2005. **102**(51): p. 18514-9.
55. Van Heuverswyn, F., et al., *Human immunodeficiency viruses: SIV infection in wild gorillas*. Nature, 2006. **444**(7116): p. 164.
56. Robertson, D.L., et al., *HIV-1 nomenclature proposal*. Science, 2000. **288**(5463): p. 55-6.
57. Woodman, Z. and C. Williamson, *HIV molecular epidemiology: transmission and adaptation to human populations*. Curr Opin HIV AIDS, 2009. **4**(4): p. 247-52.
58. Holder, M. and P.O. Lewis, *Phylogeny estimation: traditional and Bayesian approaches*. Nat Rev Genet, 2003. **4**(4): p. 275-84.
59. Posada, D., *jModelTest: phylogenetic model averaging*. Mol Biol Evol, 2008. **25**(7): p. 1253-6.
60. Tamura, K., et al., *MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0*. Mol Biol Evol, 2007. **24**(8): p. 1596-9.
61. Guindon, S., *Bayesian estimation of divergence times from large sequence alignments*. Mol Biol Evol, 2010. **27**(8): p. 1768-81.
62. Anisimova, M. and O. Gascuel, *Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative*. Syst Biol, 2006. **55**(4): p. 539-52.

63. Gray, R.H., et al., *Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda*. Lancet, 2001. **357**(9263): p. 1149-53.
64. Cohen, M.S., et al., *Prevention of HIV-1 infection with early antiretroviral therapy*. N Engl J Med, 2011. **365**(6): p. 493-505.
65. Fang, C.T., et al., *Decreased HIV transmission after a policy of providing free access to highly active antiretroviral therapy in Taiwan*. J Infect Dis, 2004. **190**(5): p. 879-85.
66. Bunnell, R., et al., *Changes in sexual behavior and risk of HIV transmission after antiretroviral therapy and prevention interventions in rural Uganda*. AIDS, 2006. **20**(1): p. 85-92.
67. Zhu, T., et al., *Genotypic and phenotypic characterization of HIV-1 patients with primary infection*. Science, 1993. **261**(5125): p. 1179-81.
68. Zhang, L.Q., et al., *Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection*. J Virol, 1993. **67**(6): p. 3345-56.
69. Fischer, W., et al., *Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing*. PLoS One, 2010. **5**(8): p. e12303.
70. Keele, B.F., et al., *Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection*. Proc Natl Acad Sci U S A, 2008. **105**(21): p. 7552-7.
71. Kearney, M., et al., *Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals*. J Virol, 2009. **83**(6): p. 2715-27.
72. Herbeck, J.T., et al., *Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes*. J Virol, 2011. **85**(15): p. 7523-34.
73. Salazar-Gonzalez, J.F., et al., *Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection*. J Exp Med, 2009. **206**(6): p. 1273-89.
74. Masharsky, A.E., et al., *A substantial transmission bottleneck among newly and recently HIV-1-infected injection drug users in St Petersburg, Russia*. J Infect Dis, 2010. **201**(11): p. 1697-702.
75. Nowak, P., et al., *The selection and evolution of viral quasispecies in HIV-1 infected children*. HIV Med, 2002. **3**(1): p. 1-11.
76. Kouyos, R.D., et al., *Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection*. Clin Infect Dis, 2011. **52**(4): p. 532-9.
77. Schuitemaker, H., A.B. van 't Wout, and P. Lusso, *Clinical significance of HIV-1 coreceptor usage*. J Transl Med, 2011. **9 Suppl 1**: p. S5.
78. Shankarappa, R., et al., *Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection*. J Virol, 1999. **73**(12): p. 10489-502.
79. Williamson, S., et al., *A statistical characterization of consistent patterns of human immunodeficiency virus evolution within infected patients*. Mol Biol Evol, 2005. **22**(3): p. 456-68.
80. Lee, H.Y., et al., *Modeling sequence evolution in acute HIV-1 infection*. J Theor Biol, 2009. **261**(2): p. 341-60.
81. Fiebig, E.W., et al., *Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection*. AIDS, 2003. **17**(13): p. 1871-9.
82. Hladik, F., et al., *Initial events in establishing vaginal entry and infection by human immunodeficiency virus type-1*. Immunity, 2007. **26**(2): p. 257-70.
83. Boggiano, C. and D.R. Littman, *HIV's vagina travelogue*. Immunity, 2007. **26**(2): p. 145-7.
84. Mattapallil, J.J., et al., *Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection*. Nature, 2005. **434**(7037): p. 1093-7.
85. Li, Q., et al., *Peak SIV replication in resting memory CD4+ T cells depletes gut lamina propria CD4+ T cells*. Nature, 2005. **434**(7037): p. 1148-52.
86. Arthos, J., et al., *HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells*. Nat Immunol, 2008. **9**(3): p. 301-9.
87. McMichael, A.J., et al., *The immune response during acute HIV-1 infection: clues for vaccine development*. Nat Rev Immunol, 2010. **10**(1): p. 11-23.
88. Mellors, J.W., et al., *Prognosis in HIV-1 infection predicted by the quantity of virus in plasma*. Science, 1996. **272**(5265): p. 1167-70.
89. Fraser, C., et al., *Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis*. Proc Natl Acad Sci U S A, 2007. **104**(44): p. 17441-6.
90. Lifson, A.R., et al., *Long-term human immunodeficiency virus infection in asymptomatic homosexual and bisexual men with normal CD4+ lymphocyte counts: immunologic and virologic characteristics*. J Infect Dis, 1991. **163**(5): p. 959-65.
91. O'Brien, S.J. and G.W. Nelson, *Human genes that limit AIDS*. Nat Genet, 2004. **36**(6): p. 565-74.
92. Hutter, G., et al., *Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation*. N Engl J Med, 2009. **360**(7): p. 692-8.

93. Piacentini, L., et al., *Genetic correlates of protection against HIV infection: the ally within*. J Intern Med, 2009. **265**(1): p. 110-24.
94. Goonetilleke, N., et al., *The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection*. J Exp Med, 2009. **206**(6): p. 1253-72.
95. Streeck, H., et al., *Human immunodeficiency virus type 1-specific CD8+ T-cell responses during primary infection are major determinants of the viral set point and loss of CD4+ T cells*. J Virol, 2009. **83**(15): p. 7641-8.
96. Tomaras, G.D., et al., *Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia*. J Virol, 2008. **82**(24): p. 12449-63.
97. Albert, J., et al., *Rapid development of isolate-specific neutralizing antibodies after primary HIV-1 infection and consequent emergence of virus variants which resist neutralization by autologous sera*. AIDS, 1990. **4**(2): p. 107-12.
98. Richman, D.D., et al., *Rapid evolution of the neutralizing antibody response to HIV type 1 infection*. Proc Natl Acad Sci U S A, 2003. **100**(7): p. 4144-9.
99. Wei, X., et al., *Antibody neutralization and escape by HIV-1*. Nature, 2003. **422**(6929): p. 307-12.
100. Dhillon, A.K., et al., *Dissecting the neutralizing antibody specificities of broadly neutralizing sera from human immunodeficiency virus type 1-infected donors*. J Virol, 2007. **81**(12): p. 6548-62.
101. Feng, Y., et al., *HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor*. Science, 1996. **272**(5263): p. 872-7.
102. Choe, H., et al., *The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates*. Cell, 1996. **85**(7): p. 1135-48.
103. Deng, H., et al., *Identification of a major co-receptor for primary isolates of HIV-1*. Nature, 1996. **381**(6584): p. 661-6.
104. Doranz, B.J., et al., *A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors*. Cell, 1996. **85**(7): p. 1149-58.
105. Dragic, T., et al., *HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5*. Nature, 1996. **381**(6584): p. 667-73.
106. Lusso, P., *HIV and the chemokine system: 10 years later*. EMBO J, 2006. **25**(3): p. 447-56.
107. Chalmet, K., et al., *Presence of CXCR4-using HIV-1 in patients with recently diagnosed infection: correlates and evidence for transmission*. J Infect Dis, 2012. **205**(2): p. 174-84.
108. Grivel, J.C., R.J. Shattock, and L.B. Margolis, *Selective transmission of R5 HIV-1 variants: where is the gatekeeper?* J Transl Med, 2011. **9 Suppl 1**: p. S6.
109. Hedskog, C., M. Mild, and J. Albert, *Transmission of the X4 phenotype of HIV-1: is there evidence against the "random transmission" hypothesis?* J Infect Dis, 2012. **205**(2): p. 163-5.
110. Karlsson, A., et al., *MT-2 cell tropism as prognostic marker for disease progression in human immunodeficiency virus type 1 infection*. J Clin Microbiol, 1994. **32**(2): p. 364-70.
111. Koot, M., et al., *Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS*. Ann Intern Med, 1993. **118**(9): p. 681-8.
112. Esbjornsson, J., et al., *Frequent CXCR4 tropism of HIV-1 subtype A and CRF02_AG during late-stage disease--indication of an evolving epidemic in West Africa*. Retrovirology, 2010. **7**: p. 23.
113. Connor, R.I., et al., *Change in coreceptor use correlates with disease progression in HIV-1--infected individuals*. J Exp Med, 1997. **185**(4): p. 621-8.
114. Scarlatti, G., et al., *In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression*. Nat Med, 1997. **3**(11): p. 1259-65.
115. Asjo, B., et al., *Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection*. Lancet, 1986. **2**(8508): p. 660-2.
116. Fenyo, E.M., et al., *Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates*. J Virol, 1988. **62**(11): p. 4414-9.
117. Cheng-Mayer, C., et al., *Biologic features of HIV-1 that correlate with virulence in the host*. Science, 1988. **240**(4848): p. 80-2.
118. Schuitemaker, H., et al., *Monocytotropic human immunodeficiency virus type 1 (HIV-1) variants detectable in all stages of HIV-1 infection lack T-cell line tropism and syncytium-inducing ability in primary T-cell culture*. J Virol, 1991. **65**(1): p. 356-63.
119. Tersmette, M., et al., *Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex*. J Virol, 1988. **62**(6): p. 2026-32.
120. Tersmette, M., et al., *Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: studies on sequential HIV isolates*. J Virol, 1989. **63**(5): p. 2118-25.

121. Moore, J.P., et al., *The CCR5 and CXCR4 coreceptors--central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection*. AIDS Res Hum Retroviruses, 2004. **20**(1): p. 111-26.
122. Baggiolini, M., B. Dewald, and B. Moser, *Human chemokines: an update*. Annu Rev Immunol, 1997. **15**: p. 675-705.
123. Rollins, B.J., *Chemokines*. Blood, 1997. **90**(3): p. 909-28.
124. Amara, A., et al., *HIV coreceptor downregulation as antiviral principle: SDF-1alpha-dependent internalization of the chemokine receptor CXCR4 contributes to inhibition of HIV replication*. J Exp Med, 1997. **186**(1): p. 139-46.
125. Oberlin, E., et al., *The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1*. Nature, 1996. **382**(6594): p. 833-5.
126. Bagnarelli, P., et al., *Host-specific modulation of the selective constraints driving human immunodeficiency virus type 1 env gene evolution*. J Virol, 1999. **73**(5): p. 3764-77.
127. Carrillo, A. and L. Ratner, *Cooperative effects of the human immunodeficiency virus type 1 envelope variable loops V1 and V3 in mediating infectivity for T cells*. J Virol, 1996. **70**(2): p. 1310-6.
128. Jansson, M., et al., *Length variation of glycoprotein 120 V2 region in relation to biological phenotypes and coreceptor usage of primary HIV type 1 isolates*. AIDS Res Hum Retroviruses, 2001. **17**(15): p. 1405-14.
129. Labrosse, B., et al., *Cooperation of the V1/V2 and V3 domains of human immunodeficiency virus type 1 gp120 for interaction with the CXCR4 receptor*. J Virol, 2001. **75**(12): p. 5457-64.
130. Edinger, A.L., et al., *Differential utilization of CCR5 by macrophage and T cell tropic simian immunodeficiency virus strains*. Proc Natl Acad Sci U S A, 1997. **94**(8): p. 4005-10.
131. Stamatos, L., M. Wiskerchen, and C. Cheng-Mayer, *Effect of major deletions in the V1 and V2 loops of a macrophage-tropic HIV type 1 isolate on viral envelope structure, cell entry, and replication*. AIDS Res Hum Retroviruses, 1998. **14**(13): p. 1129-39.
132. Shioda, T., J.A. Levy, and C. Cheng-Mayer, *Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1*. Proc Natl Acad Sci U S A, 1992. **89**(20): p. 9434-8.
133. De Jong, J.J., et al., *Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution*. J Virol, 1992. **66**(11): p. 6777-80.
134. Thordsen, I., S. Polzer, and M. Schreiber, *Infection of cells expressing CXCR4 mutants lacking N-glycosylation at the N-terminal extracellular domain is enhanced for R5X4-dualtropic human immunodeficiency virus type-1*. BMC Infect Dis, 2002. **2**: p. 31.
135. Fouchier, R.A., et al., *Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule*. J Virol, 1992. **66**(5): p. 3183-7.
136. Hu, Q., et al., *Identification of ENV determinants in V3 that influence the molecular anatomy of CCR5 utilization*. J Mol Biol, 2000. **302**(2): p. 359-75.
137. Mild, M., et al., *Differences in molecular evolution between switch (R5 to R5X4/X4-tropic) and non-switch (R5-tropic only) HIV-1 populations during infection*. Infect Genet Evol, 2010. **10**(3): p. 356-64.
138. Li, Y., et al., *Glycosylation is necessary for the correct folding of human immunodeficiency virus gp120 in CD4 binding*. J Virol, 1993. **67**(1): p. 584-8.
139. Pollakis, G., et al., *N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization*. J Biol Chem, 2001. **276**(16): p. 13433-41.
140. Derdeyn, C.A., et al., *Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission*. Science, 2004. **303**(5666): p. 2019-22.
141. Berger, E.A., et al., *A new classification for HIV-1*. Nature, 1998. **391**(6664): p. 240.
142. Koot, M., et al., *HIV-1 biological phenotype in long-term infected individuals evaluated with an MT-2 cocultivation assay*. AIDS, 1992. **6**(1): p. 49-54.
143. Whitcomb, J.M., et al., *Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism*. Antimicrob Agents Chemother, 2007. **51**(2): p. 566-75.
144. Su, Z., et al., *Response to vicriviroc in treatment-experienced subjects, as determined by an enhanced-sensitivity coreceptor tropism assay: reanalysis of AIDS clinical trials group A5211*. J Infect Dis, 2009. **200**(11): p. 1724-8.
145. Raymond, S., et al., *Development and performance of a new recombinant virus phenotypic entry assay to determine HIV-1 coreceptor usage*. J Clin Virol, 2010. **47**(2): p. 126-30.
146. de Mendoza, C., et al., *Performance of a population-based HIV-1 tropism phenotypic assay and correlation with V3 genotypic prediction tools in recent HIV-1 seroconverters*. J Acquir Immune Defic Syndr, 2008. **48**(3): p. 241-4.

147. Poveda, E., et al., *Design and validation of new genotypic tools for easy and reliable estimation of HIV tropism before using CCR5 antagonists*. J Antimicrob Chemother, 2009. **63**(5): p. 1006-10.
148. Seclen, E., et al., *High sensitivity of specific genotypic tools for detection of X4 variants in antiretroviral-experienced patients suitable to be treated with CCR5 antagonists*. J Antimicrob Chemother, 2010. **65**(7): p. 1486-92.
149. Jensen, M.A. and A.B. van 't Wout, *Predicting HIV-1 coreceptor usage with sequence analysis*. AIDS Rev, 2003. **5**(2): p. 104-12.
150. Lengauer, T., et al., *Bioinformatics prediction of HIV coreceptor usage*. Nat Biotechnol, 2007. **25**(12): p. 1407-10.
151. Vandekerckhove, L.P., et al., *European guidelines on the clinical management of HIV-1 tropism testing*. Lancet Infect Dis, 2011. **11**(5): p. 394-407.
152. Harrigan, P.R., *Screening for HIV tropism using population-based V3 genotypic analysis: a retrospective virological outcome analysis using stored plasma screening samples from MOTIVATE-1*. 5th IAS Conference on HIV Pathogenesis and Treatment, 2009. **Abstract no. WELBA101**
153. Garrido, C., et al., *Evaluation of eight different bioinformatics tools to predict viral tropism in different human immunodeficiency virus type 1 subtypes*. J Clin Microbiol, 2008. **46**(3): p. 887-91.
154. Raymond, S., et al., *Frequency of CXCR4-using viruses in primary HIV-1 infections using ultra-deep pyrosequencing*. AIDS, 2011. **25**(13): p. 1668-70.
155. Furman, P.A. and D.W. Barry, *Spectrum of antiviral activity and mechanism of action of zidovudine. An overview*. Am J Med, 1988. **85**(2A): p. 176-81.
156. Richman, D.D., *HIV chemotherapy*. Nature, 2001. **410**(6831): p. 995-1001.
157. Collier, A.C., et al., *Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine*. AIDS Clinical Trials Group. N Engl J Med, 1996. **334**(16): p. 1011-7.
158. D'Aquila, R.T., et al., *Nevirapine, zidovudine, and didanosine compared with zidovudine and didanosine in patients with HIV-1 infection. A randomized, double-blind, placebo-controlled trial*. National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group Protocol 241 Investigators. Ann Intern Med, 1996. **124**(12): p. 1019-30.
159. Staszewski, S., et al., *Virological and immunological analysis of a triple combination pilot study with zidovudine, lamivudine and zalcitabine in HIV-1-infected patients*. AIDS, 1996. **10**(5): p. F1-7.
160. Sharkey, M., et al., *In vivo evidence for instability of episomal human immunodeficiency virus type 1 cDNA*. J Virol, 2005. **79**(8): p. 5203-10.
161. Bailey, J.R., et al., *Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells*. J Virol, 2006. **80**(13): p. 6441-57.
162. Kieffer, T.L., et al., *Genotypic analysis of HIV-1 drug resistance at the limit of detection: virus production without evolution in treated adults with undetectable HIV loads*. J Infect Dis, 2004. **189**(8): p. 1452-65.
163. Nettles, R.E., et al., *Intermittent HIV-1 viremia (Blips) and drug resistance in patients receiving HAART*. JAMA, 2005. **293**(7): p. 817-29.
164. Maldarelli, F., et al., *ART suppresses plasma HIV-1 RNA to a stable set point predicted by pretherapy viremia*. PLoS Pathog, 2007. **3**(4): p. e46.
165. Butler, S.L., E.P. Johnson, and F.D. Bushman, *Human immunodeficiency virus cDNA metabolism: notable stability of two-long terminal repeat circles*. J Virol, 2002. **76**(8): p. 3739-47.
166. Pierson, T.C., et al., *Intrinsic stability of episomal circles formed during human immunodeficiency virus type 1 replication*. J Virol, 2002. **76**(8): p. 4138-44.
167. Shen, L. and R.F. Siliciano, *Viral reservoirs, residual viremia, and the potential of highly active antiretroviral therapy to eradicate HIV infection*. J Allergy Clin Immunol, 2008. **122**(1): p. 22-8.
168. St Clair, M.H., et al., *3'-Azido-3'-deoxythymidine triphosphate as an inhibitor and substrate of purified human immunodeficiency virus reverse transcriptase*. Antimicrob Agents Chemother, 1987. **31**(12): p. 1972-7.
169. Kohlstaedt, L.A., et al., *Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor*. Science, 1992. **256**(5065): p. 1783-90.
170. Kempf, D.J., et al., *Pharmacokinetic enhancement of inhibitors of the human immunodeficiency virus protease by coadministration with ritonavir*. Antimicrob Agents Chemother, 1997. **41**(3): p. 654-60.
171. Wild, C.T., et al., *Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection*. Proc Natl Acad Sci U S A, 1994. **91**(21): p. 9770-4.
172. Matos, P.M., M.A. Castanho, and N.C. Santos, *HIV-1 fusion inhibitor peptides enfuvirtide and T-1249 interact with erythrocyte and lymphocyte membranes*. PLoS One, 2010. **5**(3): p. e9830.
173. Palani, A. and J.R. Tagat, *Discovery and development of small-molecule chemokine coreceptor CCR5 antagonists*. J Med Chem, 2006. **49**(10): p. 2851-7.

174. Dorris, P., et al., *Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity*. *Antimicrob Agents Chemother*, 2005. **49**(11): p. 4721-32.
175. Espeseth, A.S., et al., *HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase*. *Proc Natl Acad Sci U S A*, 2000. **97**(21): p. 11244-9.
176. Hazuda, D.J., et al., *A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase*. *Proc Natl Acad Sci U S A*, 2004. **101**(31): p. 11233-8.
177. Finzi, D., et al., *Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy*. *Science*, 1997. **278**(5341): p. 1295-300.
178. Zhang, Z., et al., *Sexual transmission and propagation of SIV and HIV in resting and activated CD4+ T cells*. *Science*, 1999. **286**(5443): p. 1353-7.
179. Smith, B.A., et al., *Persistence of infectious HIV on follicular dendritic cells*. *J Immunol*, 2001. **166**(1): p. 690-6.
180. Popov, S., et al., *Long-term productive human immunodeficiency virus infection of CD1a-sorted myeloid dendritic cells*. *J Virol*, 2005. **79**(1): p. 602-8.
181. Kalter, D.C., et al., *Epidermal Langerhans cells are not principal reservoirs of virus in HIV disease*. *J Immunol*, 1991. **146**(10): p. 3396-404.
182. Siliciano, J.D., et al., *Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells*. *Nat Med*, 2003. **9**(6): p. 727-8.
183. *Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents*. <http://aidsinfo.nih.gov>, 2011.
184. *The European AIDS Clinical Society*. www.europeanaidsclinicalsociety.org, 2011.
185. *The Swedish guidelines for antiviral therapy* www.smittskyddsinstitutet.se/rav, 2011.
186. Wensing, A.M. and C.A. Boucher, *Worldwide transmission of drug-resistant HIV*. *AIDS Rev*, 2003. **5**(3): p. 140-55.
187. Boucher, C.A., et al., *Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects*. *J Infect Dis*, 1992. **165**(1): p. 105-10.
188. Gardner, E.M., et al., *Antiretroviral medication adherence and the development of class-specific antiretroviral resistance*. *AIDS*, 2009. **23**(9): p. 1035-46.
189. Llibre, J.M., J.M. Schapiro, and B. Clotet, *Clinical implications of genotypic resistance to the newer antiretroviral drugs in HIV-1-infected patients with virological failure*. *Clin Infect Dis*, 2010. **50**(6): p. 872-81.
190. Hogg, R.S., et al., *Emergence of drug resistance is associated with an increased risk of death among patients first starting HAART*. *PLoS Med*, 2006. **3**(9): p. e356.
191. Johnson, V.A., et al., *Update of the drug resistance mutations in HIV-1: December 2010*. *Top HIV Med*, 2010. **18**(5): p. 156-63.
192. Schinazi, R.F., et al., *Characterization of human immunodeficiency viruses resistant to oxathiolane-cytosine nucleosides*. *Antimicrob Agents Chemother*, 1993. **37**(4): p. 875-81.
193. Quan, Y., et al., *Endogenous reverse transcription assays reveal high-level resistance to the triphosphate of (-)-2'-dideoxy-3'-thiacytidine by mutated M184V human immunodeficiency virus type 1*. *J Virol*, 1996. **70**(8): p. 5642-5.
194. Meyer, P.R., et al., *Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism*. *Proc Natl Acad Sci U S A*, 1998. **95**(23): p. 13471-6.
195. Arion, D., et al., *Phenotypic mechanism of HIV-1 resistance to 3'-azido-3'-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase*. *Biochemistry*, 1998. **37**(45): p. 15908-17.
196. Larder, B.A. and S.D. Kemp, *Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT)*. *Science*, 1989. **246**(4934): p. 1155-8.
197. Bacheler, L.T., et al., *Human immunodeficiency virus type 1 mutations selected in patients failing efavirenz combination therapy*. *Antimicrob Agents Chemother*, 2000. **44**(9): p. 2475-84.
198. Demeter, L.M., et al., *Delavirdine susceptibilities and associated reverse transcriptase mutations in human immunodeficiency virus type 1 isolates from patients in a phase I/II trial of delavirdine monotherapy (ACTG 260)*. *Antimicrob Agents Chemother*, 2000. **44**(3): p. 794-7.
199. Deeks, S.G., *International perspectives on antiretroviral resistance. Nucleoside reverse transcriptase inhibitor resistance*. *J Acquir Immune Defic Syndr*, 2001. **26 Suppl 1**: p. S25-33.
200. Dykes, C., et al., *Impact of clinical reverse transcriptase sequences on the replication capacity of HIV-1 drug-resistant mutants*. *Virology*, 2001. **285**(2): p. 193-203.
201. Arts, E.J. and D.J. Hazuda, *HIV-1 Antiretroviral Drug Therapy*. Cold Spring Harb Perspect Med, 2012. **2**(4): p. a007161.

202. Tebit, D.M., et al., *Divergent evolution in reverse transcriptase (RT) of HIV-1 group O and M lineages: impact on structure, fitness, and sensitivity to nonnucleoside RT inhibitors*. J Virol, 2010. **84**(19): p. 9817-30.
203. Molla, A., et al., *Ordered accumulation of mutations in HIV protease confers resistance to ritonavir*. Nat Med, 1996. **2**(7): p. 760-6.
204. Westby, M., et al., *Reduced maximal inhibition in phenotypic susceptibility assays indicates that viral strains resistant to the CCR5 antagonist maraviroc utilize inhibitor-bound receptor for entry*. J Virol, 2007. **81**(5): p. 2359-71.
205. Coffin, J.M., *HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy*. Science, 1995. **267**(5197): p. 483-9.
206. Li, J.Z., 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**(13): p. 1327-35.
207. Simen, B.B., et al., *Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naïve patients significantly impact treatment outcomes*. J Infect Dis, 2009. **199**(5): p. 693-701.
208. Balduin, M., et al., *Prevalence of minor variants of HIV strains at reverse transcriptase position 103 in therapy-naïve patients and their impact on the virological failure*. J Clin Virol, 2009. **45**(1): p. 34-8.
209. Paredes, R., et al., *Pre-existing minority drug-resistant HIV-1 variants, adherence, and risk of antiretroviral treatment failure*. J Infect Dis, 2010. **201**(5): p. 662-71.
210. Johnson, J.A., et al., *Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naïve populations and associate with reduced treatment efficacy*. PLoS Med, 2008. **5**(7): p. e158.
211. Geretti, A.M., et al., *Low-frequency K103N strengthens the impact of transmitted drug resistance on virologic responses to first-line efavirenz or nevirapine-based highly active antiretroviral therapy*. J Acquir Immune Defic Syndr, 2009. **52**(5): p. 569-73.
212. Metzner, K.J., et al., *Minority quasispecies of drug-resistant HIV-1 that lead to early therapy failure in treatment-naïve and -adherent patients*. Clin Infect Dis, 2009. **48**(2): p. 239-47.
213. Metzner, K.J., et al., *Prevalence of key resistance mutations K65R, K103N, and M184V as minority HIV-1 variants in chronically HIV-1 infected, treatment-naïve patients*. J Clin Virol, 2011. **50**(2): p. 156-61.
214. Messiaen, P., et al., *Ultra-deep sequencing of HIV-1 reverse transcriptase before start of an NNRTI-based regimen in treatment-naïve patients*. Virology, 2012. **426**(1): p. 7-11.
215. Goodman, D.D., et al., *Low level of the K103N HIV-1 above a threshold is associated with virological failure in treatment-naïve individuals undergoing efavirenz-containing therapy*. AIDS, 2011. **25**(3): p. 325-33.
216. Metzner, K.J., et al., *Efficient suppression of minority drug-resistant HIV type 1 (HIV-1) variants present at primary HIV-1 infection by ritonavir-boosted protease inhibitor-containing antiretroviral therapy*. J Infect Dis, 2010. **201**(7): p. 1063-71.
217. Archer, J., et al., *The evolutionary analysis of emerging low frequency HIV-1 CXCR4 using variants through time--an ultra-deep approach*. PLoS Comput Biol, 2010. **6**(12): p. e1001022.
218. 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**: p. 1035-1039.
219. Yerly, S., et al., *Transmission of HIV-1 drug resistance in Switzerland: a 10-year molecular epidemiology survey*. AIDS, 2007. **21**(16): p. 2223-9.
220. SPREAD programme. *Transmission of drug-resistant HIV-1 in Europe remains limited to single classes*. AIDS, 2008. **22**: p. 625-635.
221. Karlsson, 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**(3): p. e33484.
222. Vercauteren, J., et al., *Transmission of drug-resistant HIV-1 is stabilizing in Europe*. J Infect Dis, 2009. **200**(10): p. 1503-8.
223. Yerly, S., et al., *Switch to unusual amino acids at codon 215 of the human immunodeficiency virus type 1 reverse transcriptase gene in seroconverters infected with zidovudine-resistant variants*. J Virol, 1998. **72**(5): p. 3520-3.
224. Gandhi, R.T., et al., *Progressive reversion of human immunodeficiency virus type 1 resistance mutations in vivo after transmission of a multiply drug-resistant virus*. Clin Infect Dis, 2003. **37**(12): p. 1693-8.
225. Little, S.J., et al., *Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection*. J Virol, 2008. **82**(11): p. 5510-8.
226. Lindstrom, A., et al., *HIV-1 transmission cluster with M41L 'singleton' mutation and decreased transmission of resistance in newly diagnosed Swedish homosexual men*. Antivir Ther, 2006. **11**(8): p. 1031-9.

227. Pinggen, M., et al., *Evolutionary pathways of transmitted drug-resistant HIV-1*. J Antimicrob Chemother, 2011. **66**(7): p. 1467-80.
228. Hedskog, C., et al., *Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing*. PLoS One, 2010. **5**(7): p. e11345.
229. Joos, B., et al., *HIV rebounds from latently infected cells, rather than from continuing low-level replication*. Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16725-30.
230. Diaz, R.S., et al., *Similar efficacy of lopinavir/ritonavir-containing regimens among clades B and F HIV-1-Infected individuals in Brazil*. J Acquir Immune Defic Syndr, 2008. **47**(3): p. 399-401.
231. Jourdain, G., et al., *Association between detection of HIV-1 DNA resistance mutations by a sensitive assay at initiation of antiretroviral therapy and virologic failure*. Clin Infect Dis, 2010. **50**(10): p. 1397-404.
232. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase*. J Mol Biol, 1975. **94**(3): p. 441-8.
233. Metzker, M.L., *Sequencing technologies - the next generation*. Nat Rev Genet, 2010. **11**(1): p. 31-46.
234. Dressman, D., et al., *Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations*. Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8817-22.
235. Fedurco, M., et al., *BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies*. Nucleic Acids Res, 2006. **34**(3): p. e22.
236. Hoffmann, C., et al., *DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations*. Nucleic Acids Res, 2007. **35**(13): p. e91.
237. Wang, G.P., et al., *HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications*. Genome Res, 2007. **17**(8): p. 1186-94.
238. Varghese, V., et al., *Minority variants associated with transmitted and acquired HIV-1 nonnucleoside reverse transcriptase inhibitor resistance: implications for the use of second-generation nonnucleoside reverse transcriptase inhibitors*. J Acquir Immune Defic Syndr, 2009. **52**(3): p. 309-15.
239. Ronaghi, M., M. Uhlen, and P. Nyren, *A sequencing method based on real-time pyrophosphate*. Science, 1998. **281**(5375): p. 363, 365.
240. Rozera, G., et al., *Archived HIV-1 minority variants detected by ultra-deep pyrosequencing in provirus may be fully replication competent*. AIDS, 2009. **23**(18): p. 2541-3.
241. Tsibris, A.M., et al., *Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo*. PLoS One, 2009. **4**(5): p. e5683.
242. Zagordi, O., et al., *Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies*. Nucleic Acids Res, 2010. **38**(21): p. 7400-9.
243. Zagordi, O., et al., *ShoRAH: estimating the genetic diversity of a mixed sample from next-generation sequencing data*. BMC Bioinformatics, 2011. **12**: p. 119.
244. Archer, J., et al., *Detection of low-frequency pretherapy chemokine (CXC motif) receptor 4 (CXCR4)-using HIV-1 with ultra-deep pyrosequencing*. AIDS, 2009. **23**(10): p. 1209-18.
245. De Grassi, A., et al., *Ultra-deep sequencing of a human ultraconserved region reveals somatic and constitutional genomic instability*. PLoS Biol, 2010. **8**(1): p. e1000275.
246. Campbell, P.J., et al., *Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing*. Nat Genet, 2008. **40**(6): p. 722-9.
247. Le, T., et al., *Low-abundance HIV drug-resistant viral variants in treatment-experienced persons correlate with historical antiretroviral use*. PLoS One, 2009. **4**(6): p. e6079.
248. Mitsuya, Y., et al., *Minority human immunodeficiency virus type 1 variants in antiretroviral-naïve persons with reverse transcriptase codon 215 revertant mutations*. J Virol, 2008. **82**(21): p. 10747-55.
249. Wang, C., et al., *Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance*. Genome Res, 2007. **17**(8): p. 1195-201.
250. Knoepfel, S.A., et al., *In-depth analysis of G-to-A hypermutation rate in HIV-1 env DNA induced by endogenous APOBEC3 proteins using massively parallel sequencing*. J Virol Methods, 2011. **171**(2): p. 329-38.
251. Poon, A.F., et al., *Phylogenetic analysis of population-based and deep sequencing data to identify coevolving sites in the nef gene of HIV-1*. Mol Biol Evol, 2010. **27**(4): p. 819-32.
252. Henn, M.R., et al., *Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection*. PLoS Pathog, 2012. **8**(3): p. e1002529.
253. Jabara, C.B., et al., *Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID*. Proc Natl Acad Sci U S A, 2011. **108**(50): p. 20166-71.
254. Kinde, I., et al., *Detection and quantification of rare mutations with massively parallel sequencing*. Proc Natl Acad Sci U S A, 2011. **108**(23): p. 9530-5.
255. Shafer, R.W., *Low-abundance drug-resistant HIV-1 variants: finding significance in an era of abundant diagnostic and therapeutic options*. J Infect Dis, 2009. **199**(5): p. 610-2.

256. Rozera, G., et al., *Massively parallel pyrosequencing highlights minority variants in the HIV-1 env quasispecies deriving from lymphomonocyte sub-populations*. *Retrovirology*, 2009. **6**: p. 15.
257. Palmer, S., 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**(1): p. 406-13.
258. Gianella, S., 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**(16): p. 8359-67.
259. Kanagawa, T., *Bias and artifacts in multitemplate polymerase chain reactions (PCR)*. *J Biosci Bioeng*, 2003. **96**(4): p. 317-23.
260. Meyerhans, A., J.P. Vartanian, and S. Wain-Hobson, *DNA recombination during PCR*. *Nucleic Acids Res*, 1990. **18**(7): p. 1687-91.
261. Salazar-Gonzalez, J.F., et al., *Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing*. *J Virol*, 2008. **82**(8): p. 3952-70.
262. Judo, M.S., A.B. Wedel, and C. Wilson, *Stimulation and suppression of PCR-mediated recombination*. *Nucleic Acids Res*, 1998. **26**(7): p. 1819-25.
263. Lahr, D.J. and L.A. Katz, *Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase*. *Biotechniques*, 2009. **47**(4): p. 857-66.
264. Diallo, K., M. Gotte, and M.A. Wainberg, *Molecular impact of the M184V mutation in human immunodeficiency virus type 1 reverse transcriptase*. *Antimicrob Agents Chemother*, 2003. **47**(11): p. 3377-83.
265. Johnson, J.A., et al., *Simple PCR assays improve the sensitivity of HIV-1 subtype B drug resistance testing and allow linking of resistance mutations*. *PLoS One*, 2007. **2**(7): p. e638.
266. Brennan, T.P., et al., *Analysis of human immunodeficiency virus type 1 viremia and provirus in resting CD4+ T cells reveals a novel source of residual viremia in patients on antiretroviral therapy*. *J Virol*, 2009. **83**(17): p. 8470-81.
267. Dinoso, J.B., et al., *Treatment intensification does not reduce residual HIV-1 viremia in patients on highly active antiretroviral therapy*. *Proc Natl Acad Sci U S A*, 2009. **106**(23): p. 9403-8.
268. Allers, K., et al., *Persistence of lamivudine-sensitive HIV-1 quasispecies in the presence of lamivudine in vitro and in vivo*. *J Acquir Immune Defic Syndr*, 2007. **44**(4): p. 377-85.
269. Barbour, J.D. and R.M. Grant, *The Clinical Implications of Reduced Viral Fitness*. *Curr Infect Dis Rep*, 2004. **6**(2): p. 151-158.
270. Deeks, S.G., *Treatment of antiretroviral-drug-resistant HIV-1 infection*. *Lancet*, 2003. **362**(9400): p. 2002-11.
271. Brenner, B.G., et al., *Persistence and fitness of multidrug-resistant human immunodeficiency virus type 1 acquired in primary infection*. *J Virol*, 2002. **76**(4): p. 1753-61.
272. Cong, M.E., W. Heneine, and J.G. Garcia-Lerma, *The fitness cost of mutations associated with human immunodeficiency virus type 1 drug resistance is modulated by mutational interactions*. *J Virol*, 2007. **81**(6): p. 3037-41.
273. Martinez-Picado, J., et al., *Antiretroviral resistance during successful therapy of HIV type 1 infection*. *Proc Natl Acad Sci U S A*, 2000. **97**(20): p. 10948-53.
274. Bunnik, E.M., et al., *Detection of inferred CCR5- and CXCR4-using HIV-1 variants and evolutionary intermediates using ultra-deep pyrosequencing*. *PLoS Pathog*, 2011. **7**(6): p. e1002106.
275. Salemi, M., et al., *Phylogenetics of HIV-1 in lymphoid and non-lymphoid tissues reveals a central role for the thymus in emergence of CXCR4-using quasispecies*. *PLoS One*, 2007. **2**(9): p. e950.
276. Rieder, P., et al., *Characterization of human immunodeficiency virus type 1 (HIV-1) diversity and tropism in 145 patients with primary HIV-1 infection*. *Clin Infect Dis*, 2011. **53**(12): p. 1271-9.
277. Troyer, R.M., et al., *Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression*. *J Virol*, 2005. **79**(14): p. 9006-18.
278. Huse, S.M., et al., *Accuracy and quality of massively parallel DNA pyrosequencing*. *Genome Biol*, 2007. **8**(7): p. R143.
279. Loman, N.J., et al., *Performance comparison of benchtop high-throughput sequencing platforms*. *Nat Biotechnol*, 2012.
280. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors*. *Nature*, 2005. **437**(7057): p. 376-80.