**Dissertation**

# Genotypic Analysis of HIV-1 Coreceptor Usage

Author
**Alexander Thielen**

**Dissertation**

zur Erlangung des Grades
des Doktors der Naturwissenschaften (Dr. rer. nat.)
der Naturwissenschaftlich–Technischen Fakultäten
der Universität des Saarlandes

Saarbrücken
2011

Tag des Kolloquiums:

Dekan:
Vorsitzender des Prüfungsausschusses:
Berichterstatter:

Beisitzer:

## Abstract

The acquired immunodeficiency syndrome (AIDS) is one of the biggest medical challenges in the world today. Its causative pathogen, the human immunodeficiency virus (HIV), is responsible for millions of deaths per year. Although about two dozen antiviral drugs are currently available, progression of the disease can only be delayed but patients cannot be cured.

In recent years, the new class of coreceptor antagonists has been added to the arsenal of antiretroviral drugs. These drugs block viral cell-entry by binding to one of the receptors the virus requires for infection of a cell. However, some HIV variants can also use another coreceptor so that coreceptor usage has to be tested before administration of the drug.

This thesis analyzes the use of statistical learning methods to infer HIV coreceptor usage from viral genotype. Improvements over existing methods are achieved by using sequence information of so far not used genomic regions, next generation sequencing technologies, and by combining different existing prediction systems.

In addition, HIV coreceptor usage prediction is analyzed with respect to clinical outcome in patients treated with coreceptor antagonists. The results demonstrate that inferring HIV coreceptor usage from viral genotype can be reliably used in daily routine.

## Kurzfassung

Die Immunschwächekrankheit AIDS ist eine der größten Herausforderungen weltweit. Das verursachende Humane Immundefizienz-Virus (HIV) ist verantwortlich für Millionen Tote jährlich. Obwohl es bereits mehr als zwei Dutzend verschiedene AIDS-Medikamente gibt, können diese den Krankheitsverlauf nur verlangsamen, die Patienten jedoch nicht heilen.

In den letzten Jahren wurde eine weitere Medikamentenklasse den bestehenden Therapieansätzen hinzugefügt: die Korezeptorantagonisten. Diese Wirkstoffe binden an Rezeptoren, die das Virus zum Eintritt in die Zelle benötigt und blockieren es somit. Allerdings gibt es auch Virusvarianten, die in der Lage sind Zellen mit Hilfe eines anderen Rezeptors zu infizieren. Daher sollte man vor Verschreibung eines Korezeptorantagonisten den Korezeptorgebrauch des Virus testen.

Diese Arbeit befasst sich mit der Bestimmung des Korezeptorgebrauchs aus dem viralen Erbgut mit Hilfe von statistischen Lernverfahren. Verbesserungen gegenüber existierenden Methoden werden erreicht in dem bisher nicht verwendete Genomregionen analysiert werden, durch den Gebrauch von neuesten Hochdurchsatz-Sequenziertechniken, sowie durch die Kombination von zwei existierenden Vorhersagesystemen.

Schließlich wird die Qualität der Korezeptorvorhersagen bezüglich klinischem Ansprechens bei Patienten untersucht, die mit Korezeptorantagonisten therapiert wurden. Die Ergebnisse zeigen, dass die Vorhersage des Korezeptorgebrauchs aus dem viralen Erbgut eine verläßliche Methode für den klinischen Alltag darstellt.

## Acknowledgments

The present work has been carried out in the Department for Computational Biology and Applied Algorithmics at the Max Planck Institute for Informatics in Saarbrücken. I would like to thank Prof. Thomas Lengauer for giving me the opportunity to work on this interesting topic and for creating an inspiring environment in which I had the freedom to follow my own ideas.

In particular, I have to thank Rolf Kaiser without whom most projects would not have been possible. He has always been sitting like a spider in its net bringing different groups of people together and setting up new projects again and again. However, his most important contribution was his Cato imitation, repeating that genotypic tropism testing works and will have the future in every possible situation.

Two other main contributors have to be singled out: Richard Harrigan and Martin Däumer. Both supported me with tons of data and were a never ending source of stimulating ideas and inspiring discussions.

A further very important person was my former officemate André Altmann who always had time for problems, discussions, or coffee breaks. Furthermore, I would also thank all other former and present members of our group, especially Ruth Christmann and Achim Büch who guided me through all technical, administrative and other storms. Special thanks go also to my predecessors Niko Beerenwinkel and Tobias Sing who set up this topic.

This work would not have been possible without wonderful collaborators: Thanks to Martin Däumer, Rolf Klein, and Bernhard Thiele (Institute of Genetics and Immunology, Kaiserslautern) for dragging me into the world of next generation sequencing and providing millions of sequences to be analyzed. I wish to thank the people at the British Columbia Center for Excellence in HIV/AIDS who worked with me: Andrew Low, Andrew Moores, Rachel McGovern, Luke Swenson, and Richard Harrigan. Thanks to the labs and members of the HIV-GRADE consortia for providing data and many interesting discussions, foremost Rolf Kaiser, Martin Obermeier, and Hauke Walter. I would like to thank Valentina Svicher from the University of Rome "Tor Vergata" as well as "Paco" Francisco Codoñer and Roger Paredes from IrsiCaixa in Barcelona. I am grateful to the pharmaceutical companies Pfizer Germany, Pfizer Global, and ViiV Healthcare for traveling grants and lots of data. Also many thanks to their (former) employees Laura Belkien, Marilyn Lewis, and Hernan Valdez who were wonderful contact persons.

I want to thank the "Cologne-girlies" Melanie Balduin, Maria Neumann-Fraune, Eva Heger, Elena Knops, Claudia Müller, Finja Schweitzer, Nadine Sichtig, Saleta Sierra, and Susanna Trapp as well as Jens Verheyen and Eugen Schülter who have become more than cooperation partners but also friends.

Thanks to Elena Knops, Finja Schweitzer, André Altmann, and Martin Däumer for proof-reading this thesis.

Last but definitively not least, I want to thank my family and friends for their constant love and support, especially my parents. I am grateful to Daniela for her love and always believing in me without I would not have been able to finish this thesis.

# Contents

# 1 Introduction

## Motivation

The Human Immunodeficiency Virus (HIV) has changed the lives of millions of people in the last few decades. As the causative agent of the Aquired Immunodeficiency Syndrome (AIDS) it is responsible for about 2 million deaths per year and more than 14 millions children orphaned.

Since its discovery tremendous efforts have been undertaken to understand and fight the disease and its associated virus. These efforts were and are not only at the research level but also politically and socially driven and mainly dedicated to prevent new infections. The development of antiretroviral drugs and their combination into highly active antiretroviral therapy (HAART) has improved life expectancy and quality of life substantially in the last years. Whereas 20 years ago, the diagnosis of an HIV infection was regarded as a death sentence the situation has changed dramatically in recent years.

Nowadays, we live in a world in which a person getting infected with HIV has almost the same life expectancy as a healthy person. Moreover, it is also possible to get children with only limited risk that these become infected with the virus, too.

In addition to an arsenal of new powerful drugs fighting the virus, also other developments have contributed to this improvement in life expectancy. An important cornerstone of modern anti-HIV therapy is the clinical diagnostic and the selection of appropriate therapies. In the beginning of the pandemic, there was no need to monitor the virus because only a small number of drugs was available and the patients had only few treatment options. Today, with more than two dozen antiretroviral drugs marketed, drug combinations can be selected more precisely for each individual patient.

With respect to this individual selection of therapies, patient-specific treatment of HIV infections has become one of the first fields of personalized medicine and has laid the ground for patient-specific treatment of other viral infections such as hepatitis B and C.

The reason for therapy failure is almost always the existence or emergence of drug-resistant viruses. To account for these one determines the virus' susceptibility to the drugs before their administration. In the past viral resistance was measured via phenotypic tests by assessing the virus ability to replicate under exposure of different concentrations of a single drug. These assays are time-consuming, difficult to handle and very expensive. Phenotypic drug resistance tests have therefore been replaced by genotypic approaches in which parts of the viral genome are sequenced and resistance inferred from genotype. Commonly, rules-based methods are used for this task but bioinformatics methods have been shown to be valuable tools, too, especially when optimal drug-combinations should be found. Their optimal choice is hampered by the high diversity and complexity of resistance-mutations leading to complicated rules.

With the introduction of coreceptor antagonists into anti-HIV therapy another topic

arose. These drugs block one of the two coreceptors HIV uses for entry into cells. In addition to the development of resistance mutations, the virus can therefore evade the drug by switching to the other receptor. Thus, HIV coreceptor usage also termed tropism is mandatory before administration of coreceptor antagonists. In general, this can be done as described before: with phenotypic and genotypic approaches. However, in comparison with drug resistance against traditional drugs, mutations leading to a switch in tropism seem to be less pronounced and the prediction of HIV coreceptor usage is more difficult. Therefore, bioinformatics methods are much more important than for the determination of resistance to other antiretroviral drugs.

## Overview

Prediction of HIV coreceptor usage from genotype works relatively well when the dealing with clonal data, i.e. when viruses are grown *in vitro* and then sequenced and phenotyped. Unfortunately, such predictions are generally much worse when tested on samples taken in clinics. Clinically derived samples differ from clonal data since they contain a population of often rather diverse viral variants each contributing to the bulk phenotype.

We can ask ourselves several questions to unveil reasons that might explain the poor performance of prediction methods of HIV coreceptor usage:

- Do we look at the right regions or proteins? Coreceptor usage prediction methods use the V3 loop of the viral envelope protein gp120 as input. This region is known to be the major determinant of HIV coreceptor usage but other factors might have an impact, too.

- Do we see the right variants? Minor populations of HIV variants could remain undetected with conventional sequencing methods which have a higher detection limit than phenotypic assays.

- Do our training data differ from the viral strains circulating in the patient? Clonal data often comes from lab strains that are cultivated over some time. The circulating viral strains could possibly have evolved differently.

- Are the results derived from phenotypic assays always correct? The quality of these assays has mainly been assessed using well-defined laboratory strains and not using clinically derived isolates.

Within this work, we want to address these and other points and try to find possible solutions for the aforementioned problems.

## Outline

This thesis is structured into six chapters followed by the bibliography and the appendix. In Chapter 2 we will introduce the disease AIDS, its epidemiology and the course of an infection. We will then give a comprehensive overview of how its causative pathogen HIV is composed, how it replicates and how it can be grouped into different types and subtypes.

Subsequently, we will first introduce how modern anti-HIV therapy works and then focus on the main topic of this work namely HIV coreceptor usage and its determination.

Chapter 3 will deal with different regions of the viral genome that are normally not used for coreceptor usage prediction. We will first analyze data from the second hypervariable loop of gp120, then focus on the second envelope protein gp41 and end this chapter with a brief analysis of other envelope regions and other viral proteins.

In Chapter 4 we will address minor populations usually undetected by standard sequencing technologies. A brief introduction into the 454-technology will be followed by a technical section in which we will discuss some methods we have developed or adapted to work with this new sequencing technology and how to present the results. We will then present a proof-of-concept study we have performed to show that massively parallel sequencing can be successfully used for reliable coreceptor usage prediction and finally describe the webservice we have established to deal with sequences generated with the 454-technology.

Chapter 5 will return to the classical problem: inferring HIV coreceptor usage from the V3 loop with standard sequencing methods. We will begin this chapter showing how the two most-widely used websystems can be combined into a new prediction method. Subsequently, we will discuss the effects of an updated phenotypic assay with better minority detection and and how our prediction system works on different subtypes. In the fourth section of this chapter, we will perform a comparison of different clonal and clinical datasets in order to assess the quality of our training datasets. Last but certainly not least, we will make the final step to daily diagnostics in which we will analyze coreceptor usage prediction with respect to treatment outcome under drug combinations containing coreceptor antagonists.

The thesis is concluded by Chapter 6 in which we will give a summary of the main achievements and provide an outlook of future advancements.

# 2 Background

In the following sections we will provide the biological background of this thesis. The first section (Section 2.1) will give a short overview of the disease AIDS after which we will subsequently go into more detail. We will introduce the causative agent of AIDS - the Human Immunodeficiency Virus (Section 2.2) describe how modern therapy targets the virus (Section 2.3), and finally focus on viral cell entry (Section 2.4), the main topic of this work. The last section will deal with HIV diagnostics, genotypic and phenotypic assays used in this realm and end with a brief introduction of different sequencing methods used herein (Section 2.5).

## 2.1 Aquired Immunodeficiency Syndrome (AIDS)

### 2.1.1 The Disease and its Epidemiology

The Aquired Immunodeficiency Syndrome (AIDS) is one of the leading causes of death in young adults globally. According to the AIDS epidemic update 2009 of the World Health Organization (WHO), 33.4 million people were estimated to be living with its causative pathogen the Human Immunodeficiency Virus (HIV), in 2008. This is an increase of more than 20% with respect to the numbers in 2000 and threefold higher than the prevalence in 1990. The rate of new infections is also very high with approximately 2.7 million in 2008 while about 2.0 million people died on AIDS-related illnesses. It is estimated that up to now, AIDS has caused the deaths of more than 25 million people worldwide.

AIDS comes by with symptoms that are very similar to the ones of other illnesses like tuberculosis. Consequently, these symptoms are often attributed to such well-known causes instead of an infection with HIV. Because the disease is not accompanied by signs or symptoms salient enough to be noticed, it is not very surprising that the infection has not been discovered until 1981 in which a number of young gay men in New York and Los Angeles were diagnosed with symptoms not usually seen in individuals with a healthy immune status (Hymes et al., 1981; for Disease Control, 1981; Gottlieb, 2006). These individuals showed a more aggressive form of Kaposi's sarcoma or a rare lung infection called *Pneumocystis carinii pneumonia*. Soon after these first reports, a number of further cases with similar symptoms were discovered leading to the conclusion that one was dealing with a new undefined disease associated with the breakdown of the body's immune system.

In 1983, researchers around Luc Montagnier and Françoise Barré-Sinoussi at the Institute Pasteur in Paris (France) were the first to isolate HIV (Barré-Sinoussi et al., 1983), a discovery for which they were awarded with the Nobel Price for Physiology or Medicine twenty-five years later in 2008. Montagnier and co-workers named the virus lymphadenopathy-associated virus (LAV) and suggested it might be the cause of AIDS but did not get much notice at that time. About a year later, in 1984, the United States

Health and Human Services Secretary Margaret Heckler announced that Robert C. Gallo of the National Cancer Institute had isolated the virus causing AIDS and called it HTLV-III. A very controversial debate including accusations about the first discovery of HIV started but it soon became obvious that both groups had isolated the same virus. The problem with the different names of HIV was resolved in 1986, when the International Committee on the Taxonomy of Viruses decided to drop both names and replaced them by a new term, Human Immunodeficiency Virus (HIV) (Coffin et al., 1986). Although they were not the first to have isolated the virus, the group of Gallo is recognized to be first to demonstrate that HIV is causing AIDS (Popovic et al., 1984).

Due to its first occurrences in gay men, the disease was also temporarily named *gay compromise syndrome* or *gay-related immune deficiency (GRID)* (Brennan and Durack, 1981). The common association of AIDS with men who have sex with men is one of the main reasons why AIDS is heavily stigmatized around the world until now. In March 2009, there were still 52 countries that had some form of HIV-specific restriction on entry, stay and residence. Among these, 23 countries deported individuals once their HIV-positive status is discovered while 5 countries denied visas for even short term stays (UNAIDS, 2009b). That this is a major problem not only in third-world countries is underlined by the fact that the United States just lifted its entry-ban in January 2010 and China at the end of April 2010.

The prevalence of HIV varies considerably between different regions over the world. Sub-Saharan Africa, accounting for two thirds of all people living with HIV and three fourths of AIDS deaths in 2007 is by far the most heavily affected region. In some of these countries (e.g. South Africa, Botswana and Swaziland), almost every fifth adult is infected with HIV. In contrast, most countries in Western and Central Europe have prevalence rates of about 0.1% (UNAIDS, 2008).

The reasons for the different infection rates in different regions are manifold. A major reason for the high prevalence in sub-Saharan Africa is certainly that the virus originated there through multiple infections from non-human primates infected with the simian immunodeficiency virus (SIV). Because this is believed to have happened already in 1931 (Korber et al., 2000), the virus had much time to spread until it was recognized about fifty years later. However, also other educational, social, and religious reasons such as gender equality and circumcision account for imbalances in infection rates between different regions (UNAIDS, 2008).

Another notable difference between different regions lies in the different primary modes of transmission. While the epidemic in sub-Saharan Africa is mainly driven by heterosexual exposure and often still due to mother-to-child transmissions, men who have sex with men are the most highly affected group in Latin America. Similarly, most countries in North America and Western and Central Europe show a clear trend towards increased transmission among men who have sex with men while the rates of new infections in another high-risk group of injecting drug users decreased in the last years. However, injecting drug use seems to remain the primary route of transmission in Eastern Europe, Asia, the Middle East, and North Africa. Such disproportions in the high risk groups also lead to the fact that in sub-Saharan Africa women account for about 60% of infections whereas in North America and in Western and Central Europe men outnumber women in both HIV prevalence and incidence by more than 2:1 (UNAIDS, 2009a).

In addition to the health of the infected people, AIDS has also a tremendous impact on societies as a whole in countries most heavily affected, both economically and demographically. In Swaziland, a high-prevalence country, the average life expectancy decreased by almost half from 65 years in 1991 to 37.4 years in 2005 (Whiteside et al., 2006). Such an increased mortality leads to a smaller skilled population and labor force affecting economic growth (Bell et al., 2003).

Despite the fact that more and better drugs have become available there is still a long way to go. According to the World Health Report in 2008 (UNAIDS, 2008), for every two people starting to take antiretroviral therapy, another five become newly infected. It is simply too expensive and the logistics effort too high to treat that many patients. Thus, the main focus in low income countries is not to treat patients but to prevent new HIV infections. While universal access to antiretrovirals is a desirable goal, it will probably not be possible in the near future without better HIV prevention strategies and a vaccine. Until then, clinical diagnostic tools to select appropriate antiretroviral therapies will become even more important and serve as a valuable factor in the fight against HIV.

### 2.1.2 Course of Infection

HIV is transmitted from one person to another through bodily fluids, mostly via sexual transmission across a mucosal surface. In addition, it can also be transmitted through contaminated blood transfusions, contaminated needles shared among intravenous drug users, and from mothers to their children during pregnancy, childbirth, and breast feeding.

After transmission the virus is transported to local lymph nodes by Langerhans cells or dendritic cells within the first week of infection. From there it can spread, with exponential expansion of the infection, to other sites, particularly to gut-associated lymphoid tissue (GALT) (McMichael, 2006). GALT is known to be the principal site of HIV replication (Arthos et al., 2008) and a reservoir of HIV which supports its persistence even during long-term suppression of plasma viremia while receiving antiretroviral therapy (Chun et al., 2008). At around day 21 after infection, virus appears in the blood with high viral loads of often more than 10 million copies per milliliter (McMichael, 2006). This is commonly referred to as the *acute phase* of HIV infection which is accompanied by a rapid depletion of about half of the $CD4^+$ T cells (Mattapallil et al., 2005) resulting in flu-like symptoms (see also Figure 2.1, weeks 3 to 9). It is assumed that this enormous viral replication and extensive $CD4^+$ memory T cell destruction leads to an irreversible damage to the immune system setting the stage for its eventual failure (Picker, 2006).

Towards the end of the acute phase, viral plasma levels drop 10- to 100-fold to a relatively stable *set point* between 2 and 6 months after infection while the number of $CD4^+$ T cells increases again (around week 9 in Figure 2.1). This rapid and spontaneous decline in viral load is thought to be caused by a limited target cell population (Phillips, 1996) after the initial exponential expansion and because of the appearance of specific immune responses (Borrow et al., 1994; Safrit et al., 1994). In general, the viral set point is a good predictor of how fast a patient will progress to develop AIDS (Mellors et al., 1996; Fauci, 1996; Lyles et al., 2000), the higher it is the faster a person will progress to disease.

Following the acute phase, there is a phase in which the virus can be controlled in absence of clinical symptoms over a long time, lasting from months to several years. In this phase,

called *asymptomatic or latent phase*, viral load only slowly increases whereas $CD4^+$ T cells slowly but continuously deplete (indicated in Figure 2.1 by the period between week 9 and year 7).

When the immune system is too weak and finally collapses, the third phase of the infection is reached: *AIDS*. It is commonly diagnosed when $CD4^+$ T cell counts fall below 200 cells per milliliter (starting at year 7 in Figure 2.1). Viral load increases sharply again and due to the exhausted immune system, opportunistic infections and malignancies can occur. These include cancers only rarely found among uninfected people, like Kaposi's sarcoma after infection with human herpes virus type-8 to pneumonia induced by *Pneumocystis jirovecy*. Most of these opportunistic infections would not bother healthy people too much. However, the immune system of people living with AIDS is too weak to handle them appropriately. Hence, if these patients do not get treatment, AIDS is succeeded by death from such opportunistic infections.

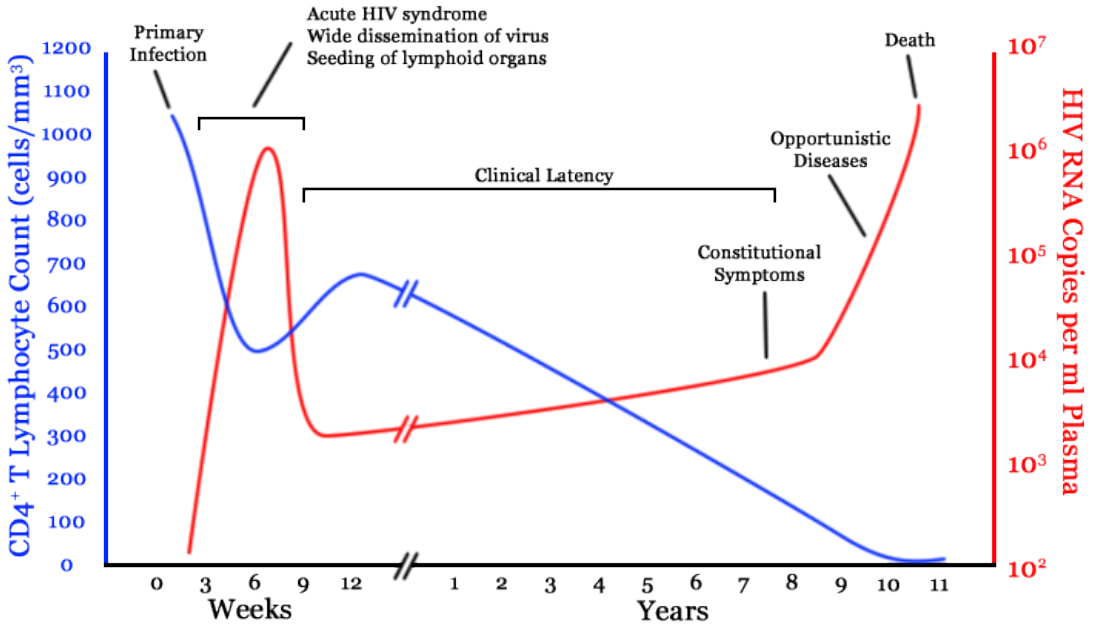


Figure 2.1: The different phases of progression to AIDS. Figure taken from Wikipedia (http://commons.wikimedia.org/wiki/File:Hiv-timecourse.png), originally published in (Pantaleo et al., 1993).

Because the symptoms of AIDS can be very diverse there is still no clear definition available for the disease. Mostly the diagnosis is referred to the 1993 Revised Classification System for HIV Infection and Expanded Surveillance Case Definition for AIDS Among Adolescents and Adults by the US Centers of Disease Control and Prevention. According to this system, the disease is diagnosed whenever an HIV-positive person has a $CD4^+$ cell count below 200 cells per microliter blood, or when the percentage of $CD4^+$ cell counts with respect to all lymphocytes is below 14%, or when the patient is diagnosed with one or more of 25 AIDS-defining illnesses. These include various opportunistic infections, brain and nerve diseases, certain cancers and the wasting syndrome. The diagnosis still stands in case that after treatment these requirements do not hold anymore.

Infection with HIV induces different responses in humans. About 10% of infected adults

are *fast progressors* showing AIDS-related symptoms within the first 2 to 3 years of infection. In approximately 40% of infected adults, progression is observed within 10 years, while 10% to 17%, commonly referred to as *long-term non-progressors*, do not progress to AIDS for more than 20 years. Some of these, termed *elite or natural controllers* or *elite surpressors*, even show no evidence of progression to disease at all and harbor plasma viremia below the limit of detection (50 copies per milliliter) (Saksena et al., 2007). These differences in progression are poorly understood, so far. They might be due to different stimulations of the immune system by infections with other pathogens as well as to viral fitness, altered cell tropism (Kupfer et al., 1998) or differences in the immune system itself (e.g. different antibody repertoires or different T-cell recognition). Recently, several studies started recruiting long-term non-progressors and elite controllers for genome-wide association studies to unveil common factors suppressing the virus naturally (Study, 2006; Fellay et al., 2007).

## 2.2 Human Immunodeficiency Virus Type 1

The Human Immunodeficiency Virus is a lentivirus belonging to the family of retroviruses. Up to now, two species of HIV are known to exist: HIV-1 and HIV-2. They were introduced into the human population by multiple cross-species transmissions from Simian Immunodeficiency Virus (SIV) infected non-human primates. HIV-1 is believed to have evolved from *SIVcpz*, a SIV-variant present in chimpanzees while HIV-2 originates from a variant (*SIVsmm*) present in sooty mangabeys (de Silva et al., 2008).

Both HIV-1 and HIV-2 are very similar in terms of transmission routes and disease symptoms (Clavel et al., 1987). However, while HIV-1 is the main driving force of the pandemic, HIV-2 is largely restricted to a very small region of West Africa (Horsburgh and Holmberg, 1988). One reason for the disproportion might be that HIV-2 has a lower transmission efficiency. A good indicator for this phenomenon turn out to be the numbers of vertical (mother-to-child) transmissions in absence of antiretroviral therapy. These are less than 4% for HIV-2 infected breast-feeding mothers, compared to up to 24.4% for HIV-1 in the same population (O'Donovan et al., 2000). For a long time, it was also assumed that immunodeficiency develops more slowly in HIV-2 infected patients (Ancelle et al., 1987). However this has been refuted by the observation that the median rate of $CD4^+$ T-cell decline per year is the same in both types of infections when matched for baseline plasma viral load (Gottlieb et al., 2002). On the other hand, most asymptomatic HIV-2 patients have undetectable baseline plasma RNA (Ariyoshi et al., 2000; Berry et al., 1998) which is associated with low or no risk of progression as in the case of HIV-1 (Ariyoshi et al., 2000; Hansmann et al., 2005).

The focus of this work will be on HIV-1. This is not so much due to differences between prevalences of these viruses but mainly because of other reasons: there is much less data available to study HIV-2 and additionally, HIV-2 uses a wide range of coreceptors instead of the more focused pattern seen in HIV-1.

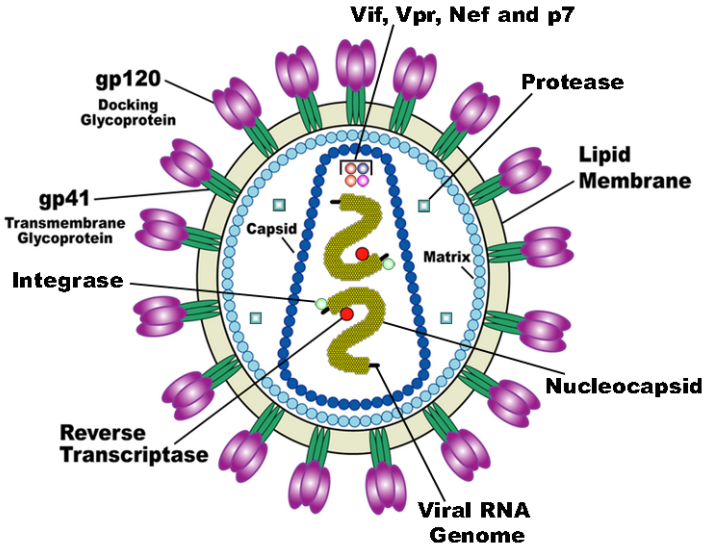


Figure 2.2: Diagram of an HIV virion. Figure taken from Wikipedia (http://commons.wikimedia.org/wiki/File:HIV_Virion-en.png).

### 2.2.1 Genes and Structure

HIV-1 virions are spherical entities with a diameter of around 100-120nm. Each particle contains two copies of single stranded RNA as well as the viral enzymes protease, reverse transcriptase and integrase in its center. Each single stranded RNA has a length of about 9.7kb and is tightly bound to nucleocapsid protein p6 and p7. The core is protected by a cone-shaped capsid comprising approximately 250 hexamers and exactly 12 pentamers of the viral p24 (CA) protein (Pornillos et al., 2009). The viral cone itself is surrounded by a spherical matrix comprised of p17 (matrix, MA) proteins which are enclosed by the viral envelope, a phospholipid bilayer. Attached to the envelope are 72 spikes composed of three copies of the viral envelope protein gp120 and the viral transmembrane protein gp41, respectively (Gelderblom, 1991). A schematic illustration of the viral structure is shown in Figure 2.2.

The single stranded RNA encodes for 15 viral proteins in total, partly in overlapping reading frames (Frankel and Young, 1998). They are synthesized from nine genes: three genes that encode for polyprotein precursors (Gag, Pol, Env) and six accessory genes with regulatory and auxiliary functions: Vif, Vpr, Vpu, Tat, Rev, and Nef. The Gag (group-specific antigen) precursor is cleaved by the viral protease into the mature Gag proteins encapsidating the viral genome: matrix (MA, p17), capsid (CA, p24), nucleocapsid (NC, p7), and p6 (Freed, 2001). Additionally, the spacer peptides 1 (SP1, p2) and 2 (SP2, p1) are encoded by the gag-gene. The Pol (polymerase) polyprotein is only expressed together with Gag as the Gag-Pol protein precursor. However, Pol is encoded in a different reading frame so that it is only synthesized in case of a -1 frameshifting event. This happens at a frequency of 5% to 10% due to a mRNA secondary structure downstream of the shift site (Parkin et al., 1992; Shehu-Xhilaga et al., 2001). The Gag-Pol precursor is cleaved by the viral protease into the viral enzymes protease (PR), reverse transcriptase (RT), and integrase (IN). The latter two transcribe the viral single-stranded RNA genome back

into double-stranded DNA and integrate it into the host genome. The Env (envelope) gene codes for the glycoprotein precursor gp160 which is cleaved by a cellular protease into the two subunits of the viral spike, the mature surface glycoprotein gp120 and the transmembrane glycoprotein gp41. The accessory genes encode for transactivators (Tat, Rev, Vpr) and other regulatory proteins (Vif, Nef, Vpu). Most of the regulatory proteins are not necessarily for viral replication *in vitro* but are important for efficient production of viruses and countering defense mechanisms of the host, *in vivo*.

The genomic structure of HIV-1 is depicted in Figure 2.3.

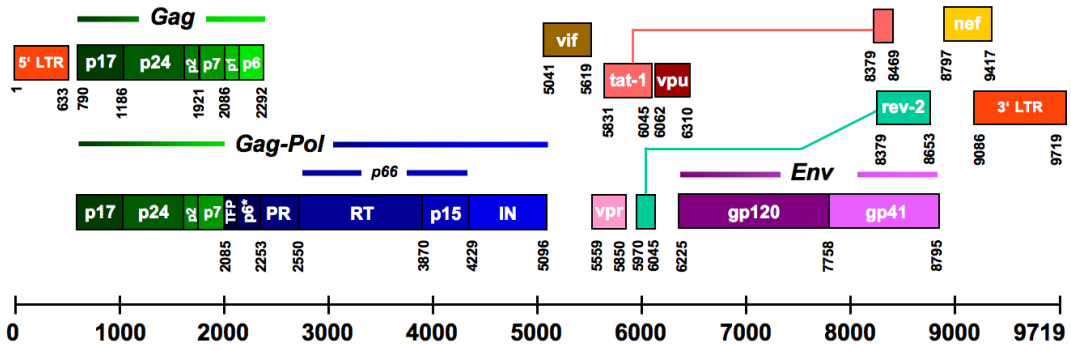


Figure 2.3: Organization of the HIV-1 genome. Figure kindly provided by Saleta Sierra and Elena Knops.

### 2.2.2 HIV-1 Replication Cycle

The life cycle of HIV-1 is a complex process that can be regarded as an ordered series of events, although some of them can occur simultaneously (Freed, 2001). The virus has an estimated turnaround time of about 1.5 days which can vary based on several factors, including target cell type and activation status. HIV replication can be divided into two overall phases: *early* and *late*. The early phase comprises the steps until the viral genome is integrated into the host genome while the late phase is characterized by the production of new virus particles: expression, assembly, and release. In the following, the individual steps of the viral replication cycle are described.

**Viral Entry** For replication HIV-1 requires a host cell expressing the CD4-receptor on its cell surface. Therefore, the virus mainly targets $CD4^+$ T lymphocytes, macrophages, microglia, and dendritic cells. It attaches to the target cell by its surface envelope glycoprotein gp120 binding to a CD4-receptor on the host-cell (Dalgleish et al., 1984). Through this process, conformational changes in the gp120 protein are induced facilitating the exposure of a second binding site for a chemokine receptor (termed coreceptor). *In vivo*, this coreceptor is either the chemokine CC receptor 5 (CCR5) or the chemokine CXC receptor 4 (CXCR4) (Berger et al., 1999; Wyatt and Sodroski, 1998). Binding to such a coreceptor activates the transmembrane glycoprotein gp41 which reorients parallel to viral and cellular membranes, allowing its N-terminal fusion peptide to insert into the target membrane. The viral and cellular membranes are then brought into close proximity (Esté and Telenti,

2007; Wyatt and Sodroski, 1998), so that in the last step of viral entry the membranes are fused and the viral core is released into the cytoplasm of the cell.

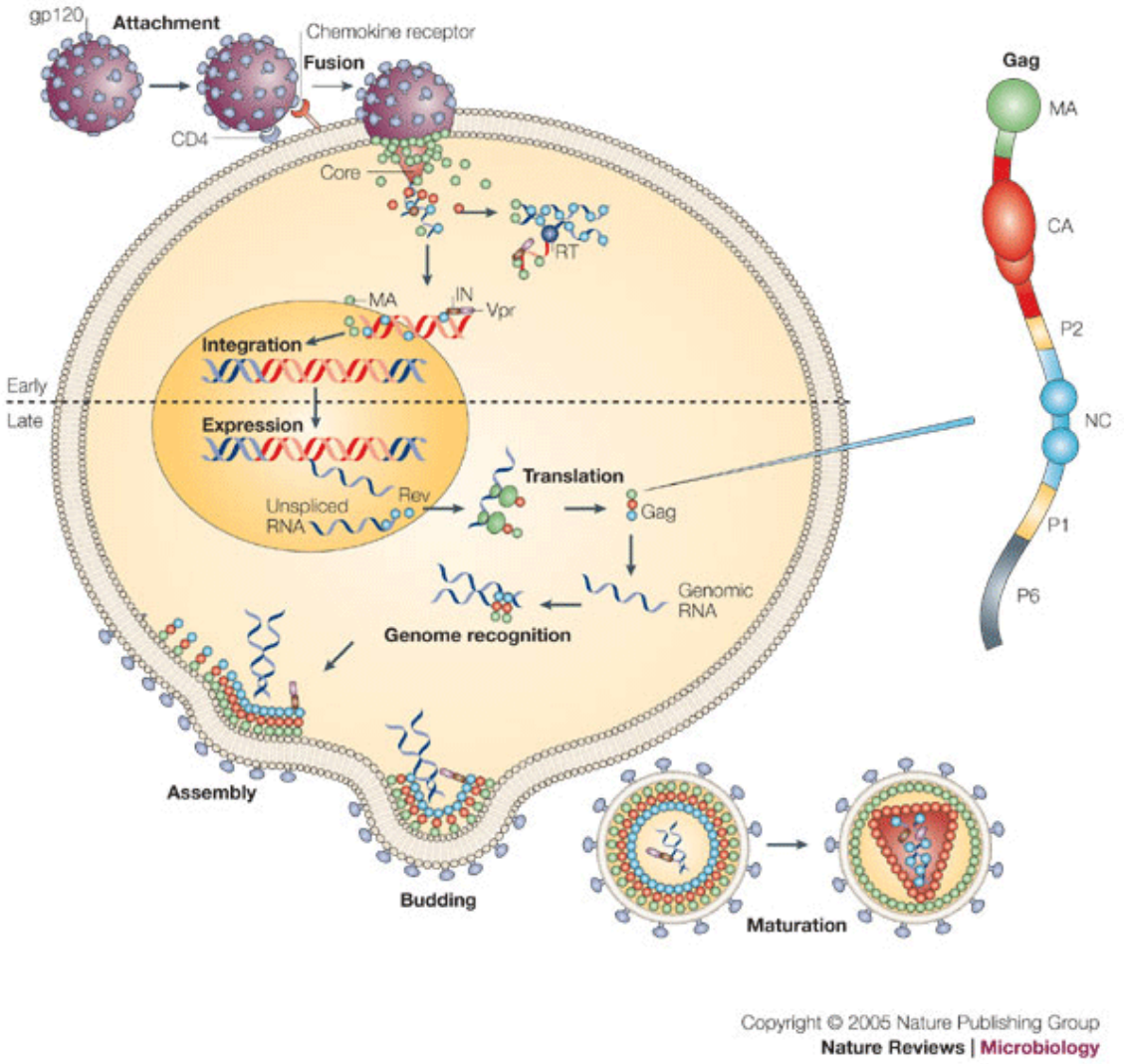


Figure 2.4: General features of the replication cycle of HIV. Viral entry, reverse transcription, nuclear import, and integration occur in the early phase of infection (upper portion of the figure). In the late phase of infection, the viral proteins are expressed, transported to the plasma membrane, assembled to mature virions and finally released. Figure derived from (D'Souza and Summers, 2005) with kind permission from Nature Publishing Group.

**Reverse Transcription, Nuclear Import, and Integration** Following the release into the cell, the viral core is *uncoated*, a process that is not completely understood yet. A large ribonucleoprotein structure called reverse transcription complex (RTC) is built up. It consists of the genomic RNA coated with nucleocapsid protein molecules and other components such as the matrix, the capsid and Vpr molecules together with the viral enzymes reverse transcriptase (RT) and integrase (IN) (Mougel et al., 2009). The reverse tran-

scriptase transcribes two single stranded copies of viral RNA into double stranded DNA (dsDNA). Since the reverse transcriptase has no proofreading activity and can also "jump" from one template to another, this process is usually error-prone and the resulting dsDNA not an exact genetic copy of the original RNA. As a consequence, there is a high frequency of genetic recombination, together with a high mutation rate ($3 \cdot 10^{-5}$ per replication cycle) leading to a highly heterogeneous population (termed *quasispecies*) (Freed, 2001; Jung et al., 2002).

Together with other cellular and viral proteins, the dsDNA forms the preintegration complex (PIC) which is transported to the nucleus (Freed, 2001). In the nucleus, the integrase being part of the PIC, catalyzes the insertion of the dsDNA into the host chromosome (Freed, 2001). The integrated dsDNA, also called the *provirus*, can then be expressed by the cell as if it was an own gene. Integration is an irreversible process marking an important point of HIV infection (Simon et al., 2006).

**Gene Expression** The integrated viral genes are transcribed by the machinery of the cell. Initially, gene expression of the viral genes is only induced at a relatively low basal level. However, this soon changes because the produced Tat (trans-activator of transcription) proteins establish a powerful positive feedback loop that increases their own expression efficiency by orders of magnitude. This, in turn, leads to very high levels of HIV-1-specific RNA and protein synthesis (Cullen, 1991; Kim and Sharp, 2001).

While Tat enhances its own expression, the viral protein Rev (regulator of virion) does quite the opposite. It regulates down its own synthesis and that of other *early class mRNAs*. The early class viral mRNAs are multiply spliced mRNAs that encode the viral regulatory proteins Tat, Nef, and Rev. In contrast, in the late phase of replication unspliced and partially spliced transcripts are generated encoding the structural proteins. The problem HIV-1 faces is that most cellular mRNAs are fully spliced before their transport out of the nucleus. The Rev protein solves this problem together with the *Rev responsive element* (RRE), a highly structured RNA element that is located in the env gene and that is present in all unspliced and partially spliced HIV-1 RNAs. Binding of approximately eight Rev proteins with one RRE results in a formation that is capable of interacting with the cellular nuclear export machinery and transporting the unspliced or partially spliced mRNAs to the cytoplasm. This process however requires a critical threshold level of Rev so that it prolongs the entry into the late phase in cells as long as they are unable to support efficient HIV-1 gene expression (Cullen, 1991; Freed, 2001).

**Budding and Maturation** The envelope polyprotein gp160 is transported to the plasma membrane via the endoplasmatic reticulum and the Golgi complex where it is subjected to extensive N-glycosylation resulting in long mannose chains. It then undergoes trimerization and is cleaved by a host protease into its gp120 and gp41 subunits (Freed, 2001; Moulard and Decroly, 2000). Both subunits are then transported to the plasma membrane where gp41 anchors the Env complex into the membrane and associates non-covalently with gp120. The precursor proteins Gag and GagPol also associate at the plasma membrane of the host where the formation of new virus particles begins.

The immature particles bud from the host cell from which they take a part of the plasma membrane as their envelope. During or shortly after budding from the cell maturation

starts. The viral protease being part of the GagPol precursor protein gets activated when the respective domains of two GagPol precursors dimerize. PR then cleaves the Gag and GagPol polyproteins into their subunits. The structural proteins of the Gag gene form the matrix, the capsid, and the nucleocapsid which rearrange to form the mature particle. Finally, only fully mature viruses are able to infect other cells.

In addition to the described steps of the replication cycle, several other actions are taken by the virus to ensure efficient replication and to defend itself against host-defense. The Nef protein, for example, is a virulence factor critical for the development of AIDS that interacts with several cellular factors to promote viral replication and immune evasion (Arhel and Kirchhoff, 2009). Its best-characterized function is the ability to reduce the number of CD4 surface molecules, promoting virus release and infectivity, and lowering the antiviral immune response. Another well-known facility of the virus is the downregulation of the major histocompatibility complex I and II molecules presenting antigens on cell surfaces. In this way, the virus can impair the recognition of HIV-infected cells by cytotoxic T cells (CTL) and T helper cells (Quaranta et al., 2009).

Another virus protein Vif (viral infectivity factor) counteracts the antiviral effects of the APOBEC3 (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide 3) proteins APOBEC3F and APOBEC3G. APOBEC proteins are involved in host innate immunity. Their antiviral activity has been shown against a number of viruses, particularly retroviruses. They induce deamination of cytidines leading to the occurrence of uridines in newly synthesized single-stranded DNA during reverse transcription and consequently G-to-A mutations in the viral genome. In absence of fully active Vif, the viral genomes are hypermutated to such an extent that they cannot produce infectious viruses anymore (Romani et al., 2009).

### 2.2.3 HIV-1 Diversity and Subtypes

One of the trademarks of HIV is its high genetic variability, posing a problem both for therapy with antiretroviral drugs as well as for the development of a protective vaccine.

As briefly mentioned before, intrapatient variability which typically manifests in the rapid development of drug resistance mutations under failing drug therapy is mainly due to the error-prone reverse transcription of the viral genome (see 2.2.1). In addition to mutations induced by the RT, interpatient variability is also characterized by different *types* of HIV (HIV-1 and -2), as well as *groups* and *subtypes* described in the following.

HIV-1 and HIV-2 are presumed to have evolved from different Simian Immunodeficiency Viruses. HIV-1 is closely related to an SIV variant (SIVcpz) present in chimpanzees (particularly Pan troglodytes troglodytes and Pan troglodytes schweinfurthii), whereas the origin of HIV-2 is probably SIVsmm which is found in high frequencies in sooty mangabey monkeys (Cercocebus atys).

Both types are further divided into different groups that are the result of multiple cross-species transmissions of the respective SIV-variants to humans. In case of HIV-1, these are the groups M (major), N (non-M, non-O), and O (outlier). Group M is distributed all over the world and is the main source of the pandemic. The groups N and O are restricted to individuals of West African origin (Rambaut et al., 2004) and are only rarely seen. Very

recently, an additional group P virus has been discovered that is closely related to SIVgor found in wild-living gorillas (Gorilla gorilla gorilla) (Plantier et al., 2009). For HIV-2, the groups A-H are known but only groups A and B have infected substantial numbers of people in West Africa (Santiago et al., 2005).

The M group of HIV-1 is further subdivided into phylogenetically associated groups of HIV-1 sequences, termed *clades* or *subtypes* (Robertson et al., 2000, see also Figure 2.5). Sequences within a subtype are more similar to each other than to sequences from another subtype. The genetic distance between two subtypes is about 25% to 35% (Butler et al., 2007), depending on the genes compared. So far, subtypes A to K are known. The different subtypes are very unevenly distributed throughout the world and often occur predominantly in certain geographic regions. For example, subtype C responsible for about half of the infections worldwide is the predominant subtype in Southern and East Africa, and in India. Subtype B on the other hand, was and is one of the driving forces for new infections in Europe and North America, and is by far the most studied subtype.

If a patient gets infected multiply with viruses from different subtypes, it can occasionally happen that two viruses recombine to a new viral strain with a mosaic genome composed of regions from both subtypes. In most cases, these recombinant viruses only occur within a single patient but in case the patient infects another person they can also spread into the population. They are then called unique recombinant forms (URFs), and if at least three representative full-length genomes have been sequenced from individuals whose infections cannot be epidemiologically linked, they are termed circulating recombinant forms (CRFs).

In fact, two of the previously determined subtypes, namely E and I, are recombinants of two or more subtypes. This is because in the beginning, subtypes were determined only by comparing the env region and not the full-length viruses. Later it became apparent that subtype E is the result of a recombination event involving the subtype A gag region and the env region of another strain putative from subtype E. However, no "pure", non-recombinant form, of subtype E has ever been found. The recombinant strain has been renamed to CRF01_AE but is sometimes still referred to as subtype E (Butler et al., 2007). Similarly, subtype I has later been found to have a complex mosaic structure of subtypes A, G, H, K and unclassified regions and has therefore been renamed to CRF06_cpx (Paraskevis et al., 2001).

Last but not least, because there is significant diversity within subtypes, sub-subtypes have been established. E.g. subtype F consists of the sister-clades F1 and F2 while subtype A is nowadays split into sub-subtypes A1–A4 (Butler et al., 2007).

Subtypes do not only differ in their sequences but of course also in the effects that these genomic differences induce. For example, subtype D seems to be more virulent than subtype A resulting in significantly faster rates of $CD4^+$ T-cell losses and more rapid disease progression (Kiwanuka et al., 2009; Baeten et al., 2007). In another study it has been shown that survival among individuals in Thailand infected with CRF01_AE appears to be similar to that reported among individuals in Africa infected with HIV-1 subtype D and significantly less than that reported among persons of similar age in high-income countries or in eastern or southern Africa (Nelson et al., 2007). An explanation for these differences might be the different preferences among subtypes for HIV-1 coreceptor usage which is associated with progression to disease (Kupfer et al., 1998). Some subtypes tend to use almost exclusively the CCR5-coreceptor while others show a higher prevalence for

CXCR4 (Tscherning et al., 1998; Nedellec et al., 2009).

It should also be noted that antiretroviral drugs are developed mainly against subtype B viruses, the most prevalent subtype in Europe and North America. This can have a negative impact on the effectiveness of these drugs for other subtypes most prevalent in the high prevalence countries of Africa and Asia (Descamps et al., 1998; Martínez-Cajas et al., 2008).

With respect to the development of an HIV vaccine, the high diversity of HIV is one of the major obstacles because such a vaccine would need to be effective against most subtypes. Thus, a vaccine would probably rely on a conserved region of the virus. Such regions are very hard to find even within a single subtype. Within the same subtype, however, HIV-1 can vary up to 20% in sequence. When comparing two subtypes the sequence diversity can rise to 38%. The challenge of such a high diversity becomes even more apparent when one compares the variability of HIV-1 within a single infected individual with the one generated over the course of a global influenza epidemic. Six years after infection, the variation of HIV-viruses within a patient is roughly the same as the one of the influenza A infections worldwide in 1996 (Walker and Burton, 2008; Weiss, 2003) (see also Figure 2.6). Since every year a new vaccine is needed against influenza one can imagine how powerful an HIV vaccine would need to be to control the global pandemic.

## 2.3 Anti-HIV Therapy

Except for a single specific patient who seems to have been cleared from HIV by therapy, there is no case known in which a person could be cured from AIDS. This patient received an allogenic stem-cell transplantation with stem-cells homozygous for a defect that renders the gene for the CCR5 coreceptor nonfunctional. Unfortunately, while the approach is quite clever, it will also remain a very rare exception. This is because only about 1% of the Caucasian population are homozygous for the CCR5 gene-defect (Novembre et al., 2005) and finding a suitable donor for a stem-cell transplation is already difficult without this additional criterion. Moreover, such a therapy is definitively not advisable because of the severe side effects arising when shutting down the human immune system for a bone marrow transplantation. Nevertheless, the case of the so-called Berlin-patient will probably be a valuable resource for the development of future gene therapies (Hütter et al., 2009).

The best way of fighting HIV would certainly be a vaccine protecting against it. After discovering that HIV causes AIDS, researchers expected to have a vaccine available in just a few years (Smith, 2003; Walker and Burton, 2008). The hunt for it, however, has become an unfortunate saga lasting now for more than 25 years (Smith, 2003, 2005, 2006). Nowadays, there is still no vaccine in sight and recent results were more frustrating than encouraging. The STEP-trial testing a promising vaccine candidate in a study containing more than 3.000 people was more or less a complete disaster. Instead of preventing HIV-1 infection or to reduce early viral level, individuals treated with the vaccine showed higher rates of infection than patients in the control group (Buchbinder et al., 2008).

Results from another recent study investigating the AIDSVAX vaccine candidate were a bit more promising but the results should interpreted with care as only a few people got infected and statistical significance was marginal and only apparent after some individuals were excluded from analysis (Rerks-Ngarm et al., 2009).

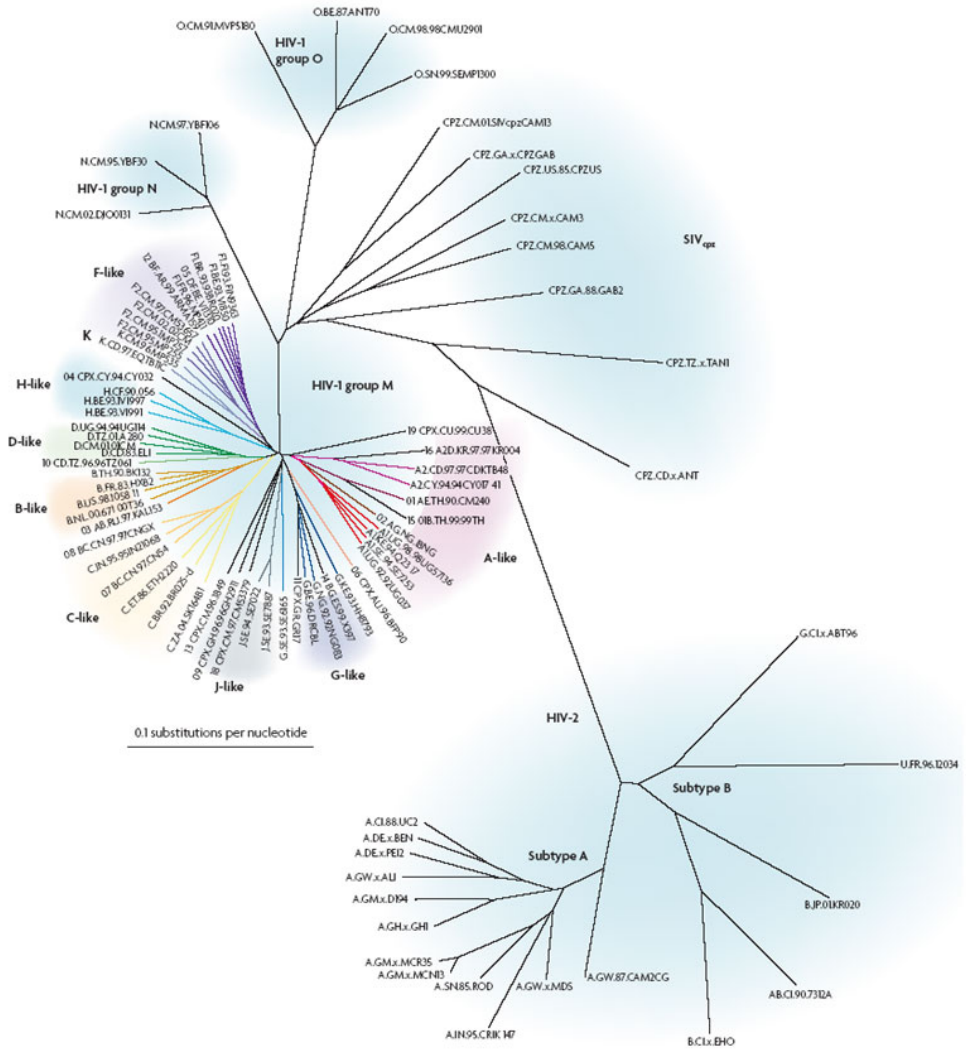


Figure 2.5: Phylogenetic tree of different human (HIV) and simian (SIV) immunodeficiency viruses. The newly identified group P and its closely related SIVgor variant as well as the HIV-2 subtypes C-H are not shown. Figure reprinted from (Ariën et al., 2007) with kind permission from Nature Publishing Group.

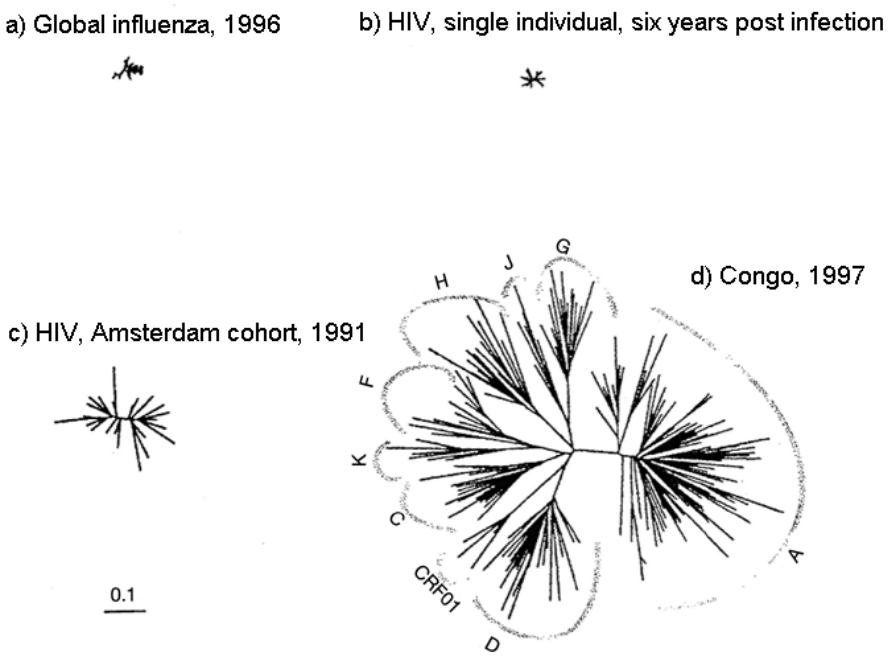


Figure 2.6: Intra- and interpatient variation of HIV in comparison with the variation of the global influenza infection. Sequence diversity is based on the envelope glycoproteins of HIV and influenza A H3 (HA1). The variation of HIV within a single patient six years after infection (b) is similar to that of worldwide influenza A in 1996 (a). For comparison also the diversity of HIV strains in the Amsterdam cohort (c) and in a set of strains with different subtypes in the Democratic Republic of Congo is shown (d). Derived and adapted from (Weiss, 2003) with kind permission from Nature Publishing Group.

Due to these reasons, HIV therapy is currently based on antiretroviral drugs. Growing understanding of the viral replication cycle has allowed the development of a whole battery of drugs targeting most steps of HIV-1 life cycle. However, so far, all these drugs can only delay the disease and prolong life. They cannot clear the virus from the body. This is because once the virus is integrated in the human genome, it is not possible, at least by current technologies available, to cut it out again. In addition, eliminating all infected cells is also not an option because there are too many in too many tissues. Antiviral therapy can suppress viral replication below the level of detection but even under a successful therapy the virus continues to replicate and mutate. This has the effect that after some time, resistant viruses are generated rendering the therapy ineffective.

The first antiretroviral drug to be used in treatment against HIV-1, azidothymidine (AZT) also known as zidovudine (ZDV), was approved by the FDA in March 1987. AZT is a nucleoside analog reverse transcriptase inhibitor that was originally designed against cancer in 1964. Although it failed to prove effective against cancer it later demonstrated potent activity against HIV (Fischl et al., 1987). The approval of AZT seemed to be a breakthrough in the fight against AIDS, but unfortunately it soon became clear that the virus is able to develop resistance against the drug during treatment (Larder and Kemp, 1989; Larder et al., 1989). The same effect has been observed for all drugs given alone (monotherapy) since then (Johnson et al., 2008).

Following the approval of AZT in 1987, more than twenty antiretroviral drugs have been approved for treating HIV by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMEA). These drugs can be classified into several groups according to the steps in the viral replication cycle with which they aim to interfere: there are protease inhibitors, reverse transcriptase inhibitors, integrase inhibitors, and entry inhibitors. In addition to these, further drugs not only from these but also from novel drug classes are in development. An overview of the different drug classes and their mechanism of action will be given in the following.

### 2.3.1 Antiretroviral Drugs

#### Protease Inhibitors

The protease is important for the virus during maturation when it cleaves the Gag and GagPol precursor polyproteins into their functional subunits. It is only active as a homodimer in which the active site is formed by a cleft between the two monomers.

The cleft has a length of about seven amino acids. Protease inhibitors (PIs) are small molecules that aim to interfere with this process by blocking the active site of the enzyme. They mimic the structure of the viral peptides that are naturally recognized thereby outcompeting them. The precursor proteins cannot be processed into the functional enzymes so that the immature viruses cannot bud from the cell and infect new cells (Wensing et al., 2010).

Because protease inhibitors interfere with the late phase of the viral replication cycle, they cannot prevent the virus from integration and consequently from establishing a persistent infection in the cells. Thus, these cells are temporarily replication-incompetent during a successful therapy but will remain potential viral distributors that can generate new viral particles in the absence of drug pressure.

HIV can react against PIs by reducing its susceptibility to the inhibitor with mutations enlarging the catalytic site of the PR. However, these resistance mutations often lead to some decrease in binding affinity to the natural substrate, too. Thus, they are usually accompanied with *accessory protease inhibitor resistance mutations* that up-regulate protease processivity to compensate for the decreased fitness associated with the major PI resistance mutations (Shafer and Schapiro, 2008; Wensing et al., 2010). Because the processing efficiency of the viral protease depends on the enzyme itself as well as on its substrate, these accessory mutations do not necessarily have to occur in the enzyme but can also be modulated by cleavage site mutations in the gag protein (Nijhuis et al., 2008; Verheyen et al., 2006).

PIs are usually very potent both in first-line therapy as well as in subsequent treatments. Due to their relatively poor bioavailability, they are mostly "boosted" with a low dose of ritonavir, the first protease inhibitor approved in the European Union. Ritonavir inhibits the hepatic and intestinal cytochrome P450 pathway mediating metabolism of most PIs. Thus, administered in combination with other PIs, it improves their bioavailability and the half-life of these drugs dramatically (Wensing et al., 2010).

The drawbacks of PIs are high costs and severe side effects. Despite suppressing the virus successfully PIs sometimes have to be replaced by other drugs.

#### Reverse Transcriptase Inhibitors

Reverse transcriptase inhibitors form the largest class of drugs against HIV, to date. Two different types of reverse transcriptase inhibitors exist, inhibiting the reverse transcription of the viral genome to double-stranded DNA in two ways: Nucleoside reverse transcriptase inhibitors (NRTIs) are deoxynucleotide-analogues that compete with their corresponding nucleotides for incorporation into the newly synthesized DNA strand. In contrast to natural deoxynucleotides, the analogues lack a 3'-hydroxyl group needed for elongation of the DNA strand. Thus, when incorporated into the strand they act as chain terminators inhibiting the critical step of reverse transcription prior to integration into the host cell genome (Cihlar and Ray, 2010; Shafer and Schapiro, 2008).

Viral resistance against NRTIs is usually achieved either by reduced susceptibility to nucleoside-analoga and enhanced incorporation of the natural nucleotides, or by enhanced removal of the chain terminating NRTIs at the 3' end by promoting a phosphorolytic reaction that leads to primer unblocking (Shafer and Schapiro, 2008).

NRTIs form the backbone of modern HAART in which at least one NRTI is commonly used. Reasons for this drug selection are the reduced pill burden, increased adherence, and better control of side effects in comparison to PIs (Cihlar and Ray, 2010).

The second drug group inhibiting the RT is formed by the non-nucleoside reverse transcriptase inhibitors (NNRTIs). They bind to a hydrophobic pocket close to the catalytic site of the RT. Binding to this part of the enzyme triggers conformational changes reducing the flexibility of the RT that is needed for viral DNA synthesis (de Béthune, 2010).

Resistance to NNRTIs occurs by mutations that lower susceptibility to the inhibitor. In general, one or two mutations provoked by therapy with a single NNRTI confer high-level cross-resistance against all NNRTIs. This is not completely true for the newest NNRTI etravirine which has specifically been designed to be susceptible in the presence of

resistance mutations by other NNRTIs (Shafer and Schapiro, 2008). In general, NNRTIs are safe and well-tolerated and therefore usually used in first-line therapy. They are also often prescribed to prevent mother-to-child transmission. The drawback of NNRTIs is the high-level of cross-resistance between the different drugs in this class often precluding use of other NNRTIs after one has failed (de Béthune, 2010).

### Integrase Inhibitors

Integrase inhibitors (INIs) are a new class of drugs that aim to prevent integration of the viral DNA into the host genome. Currently, only one approved INI is available: raltegravir (RAL). RAL preferentially inhibits the strand transfer by binding to the target DNA site of the viral integrase. Strand transfer is a process during integration in which the viral DNA 3'-ends are linked covalently to the target DNA. Interfering with this step inhibits integration and therefore also interrupts the life cycle of HIV (McColl and Chen, 2010).

Resistance to RAL is achieved by primary mutations in the proximity of the IN binding pocket decreasing the susceptibility to the drug. Secondary mutations compensate for decreased fitness induced by the primary resistance mutations. A high level of cross-resistance has been observed between resistance mutations against RAL and elvitegravir, a second strand transfer inhibitor currently being in clinical trials (Shafer and Schapiro, 2008).

Raltegravir has only recently entered clinical routine and experience with it is still limited. However it has been shown to be a potent, well tolerated drug that can be combined with existing antiretroviral drugs to provide highly efficacious treatment to both treatment experienced and treatment naïve subjects in the clinical approval studies (McColl and Chen, 2010).

### Entry Inhibitors

Entry inhibitors aim at interfering with the virus at the earliest step of its replication cycle, preventing viral cell entry. Two different groups of entry inhibitors exist, so far. Fusion inhibitors prevent fusion of the viral and host cell membranes whereas coreceptor antagonists block binding of the viral gp120 protein to the coreceptor which the virus requires for successful infection.

**Fusion Inhibitors** Enfuvirtide (ENF), formerly known as T-20, is currently the only licensed fusion inhibitor. The synthetic peptide of 36 residues is designed to mimic the amino acids 127-162 of the the heptad repeat (HR) 2 of gp41. In the process of membrane fusion, this helical region binds to another heptad repeat (HR1) forming the ectodomain of gp41. The ectodomain undergoes a conformational rearrangement resulting in a six-helix bundle structure that is thought to bring the virus and cell membranes into close proximity. This leads to fusion pore formation and membrane fusion. ENF inhibits the conformational change by replacing HR2 and blocking the formation of the helical bundle through binding to HR1 (Esté and Telenti, 2007; Menéndez-Arias, 2010).

Resistance mutations usually occur in the HR1 domain which is in absence of ENF the most conserved region in the HIV-1 envelope glycoprotein. Only little variation has been observed even among different subtypes. Nevertheless, mutations in gp41 outside HR1 and

even in the gp120 protein have been shown to reduce susceptibility against ENF (Derdeyn et al., 2000, 2001; Menéndez-Arias, 2010; Shafer and Schapiro, 2008).

The main drawbacks of Enfuvirtide are its rather high price and that is has to be administered by injection twice a day. This often causes severe side effects mainly manifesting in a number of skin irritations (Esté and Telenti, 2007), sometimes so severe that people have to stop treatment.

**Coreceptor Antagonists** Coreceptor antagonists are the newest drug class that has been added to the arsenal of antiretroviral drugs against HIV. In contrast to all other drug classes they target host, rather than viral proteins.

Targeting extracellular host proteins has two potential advantages. First, they are more accessible and drugs interfering with them do not require intracellular processing. This should minimize non-receptor-linked intracellular toxic effects such as mitochondrial toxicity (Westby and van der Ryst, 2005). Second, because the virus cannot mutate the target of these drugs, it is thought that achieving resistance to them is more difficult. The disadvantage of targeting host proteins is that this inhibits their natural function, too, which may lead to undesired side effects.

In the case of the cellular CCR5-gene, encoding for one of the two coreceptors HIV-1 uses for cell entry *in vivo*, there exists a 32-base-pair deletion allele (CCR5-$\Delta$32) preventing the functional expression of the CCR5 coreceptor. The allele is found principally in Europe and western Asia with an average frequency of approximately 10%. Approximately 1% of the Caucasian population is homozygous for CCR5-$\Delta$32 and has no major apparent deficiencies (Novembre et al., 2005; Esté and Telenti, 2007). However, individuals carrying the CCR5-$\Delta$32-deletion on both alleles have been shown to be resistant to HIV-1 infection except for very few exceptions (Samson et al., 1996b; Dean et al., 1996). These observations led to the development of CCR5-antagonists. Maraviroc (MVC), the first drug of this class has been approved for the treatment of therapy-experienced patients in 2007. It selectively binds to the CCR5-receptor thereby blocking HIV-1 entry into the cell (Westby and van der Ryst, 2010).

As anticipated, the fact that MVC targets a cellular receptor instead of a viral protein seems to complicate evolution to resistance. Unfortunately, the virus can escape MVC drug pressure by switching to an alternate coreceptor CXCR4, a process that has been observed in about two thirds of treatment failures in the clinical trails (Fätkenheuer et al., 2008). Because of the possible use of CXCR4, tropism has to be determined before administration of CCR5-antagonists (Lengauer et al., 2007).

While further CCR5-antagonists are in development, CXCR4-antagonists are currently not in clinical trials because the first investigated compounds showed liver histology changes and animal hepatotoxicity (Esté and Telenti, 2007).

### Drugs in Development

Despite of more than two-dozen antiretroviral drugs available, resistance to these compounds, drug toxicity, adverse drug-drug interactions, and poor patient adherence still require the development of new drugs (Adamson and Freed, 2010). In addition to novel agents from existing therapeutic classes (e.g. the NRTI apricitabine, the NNRTI rilpivirine,

and the INI elvitegravir are in phase 3 testing) (Esté and Cihlar, 2010), also new approaches targeting other points of the viral replication cycle are in various stages of clinical development.

Maturation inhibitors compose one of the new drug classes and have already entered clinical trials. The most advanced compound of this new class is bevirimat (BVM), which prevents cleavage of the Gag precursor protein by binding to the Gag protein and thereby blocking the viral protease. Although BVM has been demonstrated to be potent against HIV in both phase I and phase II clinical studies (Smith et al., 2007), recent data have revealed the existence of naturally occurring polymorphisms that confer reduced susceptibility to the drug. Thus, it is unclear at the moment, if the drug will be further investigated (Adamson et al., 2010; Wainberg and Albert, 2010). Nevertheless, bevirimat has shown that maturation inhibitors could be a valuable addition to the antiretroviral repertoire and further compounds of this class are already in early clinical studies.

Other novel targets and agents like Tat-TAR interaction inhibitors, anti-sense oligonucleotides, and RNase H inhibitors have shown their potential in proof-of-concept studies but are still in early stages of development. They will require a lot of further intensive research to become future antiretroviral drugs (Bhattacharya and Osman, 2009; Esté and Cihlar, 2010).

In addition to the search for new targets, research has also focused on pharmacoenhancing agents. These aim to improve the effectiveness of existing drugs and to reduce side effects of current regimens by replacing ritonavir as the only available booster (Esté and Cihlar, 2010).

### 2.3.2 Viral Resistance and Highly Active Antiretroviral Therapy

Antiretroviral drugs have been shown to reduce viral replication to levels undetectable for standard technologies. However, HIV-1 develops resistance to all of these drugs sooner or later, especially when given alone or in combination with drugs with impaired efficacy.

Drug resistance usually originates from genomic variation due to misincorporated nucleotides during reverse transcription. While mutations always occur in the viral population, resistance mutations normally only manifest when the respective viruses are still viable and when they have an selective advantage compared to the other viruses in the viral quasispecies. In case of antiretroviral drugs this advantage is obvious since non-resistant viruses are usually not able to replicate anymore. The time for the virus to become resistant and adapt to antiretroviral drugs can differ greatly and depends on several factors: If the drug has been given before, then drug resistant viruses usually emerge immediately. The reason for this is that HIV-1 integrates into the human genome and therefore "older versions" can be easily recovered from proviral depots in diverse host tissues. If there is no archived resistant variant, the time to become resistant depends primarily on the genetic barrier, the size of the viral population, and on (bad) luck. Genetic barrier is a term reflecting the difficulty to develop resistance. A high genetic barrier usually means that several mutations are needed to develop resistance while in the case of a low genetic barrier only a few (sometimes just one) mutations are required. Since every mutation occurs more or less independently by chance, the probability of evolving to resistance is of course lower with a high genetic barrier and consequently acquiring resistance takes longer on average.

| Abbr. | Generic Name | Brand Name | Manufacturer Name | FDA approval |
|---|---|---|---|---|
| | **Nucleoside Reverse Transcriptase Inhibitors** | | | |
| AZT,ZDV | zidovudine, azidothymidine | Retrovir | GlaxoSmithKline | 1987 |
| ddI | didanosine, dideoxyinosine | Videx | Bristol Myers-Squibb | 1991 |
| ddC | zalcitabine | Hivid | Hoffmann-La Roche | 1992 |
| d4T | stavudine | Zerit | Bristol Myers-Squibb | 1994 |
| 3TC | lamivudine | Epivir | GlaxoSmithKline | 1995 |
| ABC | abacavir sulfate | Ziagen | GlaxoSmithKline | 1998 |
| ddI EC | enteric coated didanosine | Videx EC | Bristol Myers-Squibb | 2000 |
| TDF | tenofovir disoproxil fumarate | Viread | Gilead | 2001 |
| FTC | emtricitabine | Emtriva | Gilead Sciences | 2003 |
| | **Nonnucleoside Reverse Transcriptase Inhibitors** | | | |
| NVP | nevirapine | Viramune | Boehringer Ingelheim | 1996 |
| DLV | delavirdine | Rescriptor | Pfizer | 1997 |
| EFV | efavirenz | Sustiva | Bristol Myers-Squibb | 1998 |
| ETV | etravirine | Intelence | Tibotec Therapeutics | 2008 |
| | **Protease Inhibitors** | | | |
| SQV | saquinavir mesylate | Invirase | Hoffmann-La Roche | 1995 |
| IDV | indinavir | Crixivan | Merck | 1996 |
| RTV | ritonavir | Norvir | Abbott Laboratories | 1996 |
| | saquinavir | Fortovase | Hoffmann-La Roche | 1997 |
| NFV | nelfinavir mesylate | Viracept | Agouron Pharmaceuticals | 1997 |
| APV | amprenavir | Agenerase | GlaxoSmithKline | 1999 |
| LPV/RTV | lopinavir and ritonavir | Kaletra | Abbott Laboratories | 2000 |
| FOS-APV | Fosamprenavir Calcium | Lexiva | GlaxoSmithKline | 2003 |
| ATV | atazanavir sulfate | Reyataz | Bristol-Myers Squibb | 2003 |
| TPV | tipranavir | Aptivus | Boehringer Ingelheim | 2005 |
| DRV | darunavir | Prezista | Tibotec, Inc. | 2006 |
| | **Fusion Inhibitors** | | | |
| ENF | enfuvirtide, T-20 | Fuzeon | Hoffmann-La Roche | 2003 |
| | **Entry Inhibitors - CCR5 coreceptor antagonists** | | | |
| MVC | maraviroc | Selzentry, Celsentri | Pfizer | 2007 |
| | **HIV integrase strand transfer inhibitors** | | | |
| RAL | raltegravir | Isentress | Merck & Co., Inc. | 2007 |
| | **Multi-class Combination Products and Combined NRTIs** | | | |
| | 3TC + ZDV | Combivir | GlaxoSmithKline | 1997 |
| | ABC + ZDV + 3TC | Trizivir | GlaxoSmithKline | 2000 |
| | ABC + 3TC | Epzicom | GlaxoSmithKline | 2004 |
| | TDF + FTC | Truvada | Gilead Sciences, Inc. | 2004 |
| | EFV + FTC + TDF | Atripla | Bristol-Myers Squibb and Gilead Sciences | 2006 |

Table 2.1: Antiretroviral drugs approved by the FDA (adapted from: http://www.fda.gov/oashi/aids/virals.html). All drug names are registered trademarks and listed in order of FDA approval within each class. Hivid and Fortovase are no longer marketed.

The larger the viral population and the higher the viral turnaround, the more possibilities the viral population has to find a resistant variant.

Chance is a major issue in this context. It might happen that the viral reverse transcriptase generates all required mutations within the first round of replication but it also might happen that (although unrealistic) all viruses simultaneously acquire a mutation that is lethal for the virus or that they never find the resistance mutations.

The extraordinary high turnover, with a total daily virion production of more than $10^{10}$ in the initial stages of the infection, coupled with a replication time of only one to three days and a high mutation rate is the main reason why under monotherapy resistant variants appear already within the first weeks (Simon and Ho, 2003).

Based on this knowledge, it was suggested to combine several drugs into a *combination therapy*. The idea was to use several drugs to increase the genetic barrier which in turn should make it more difficult to gather resistance mutations (Hirsch, 1990). In 1995, the hypothesis first proved effective in two clinical trials, Delta and ACTG175. It was demonstrated that the combination of two drugs, namely AZT with ddI or ddC, is more effective than medication with AZT alone leading to the fact that dual therapy comprising two NRTIs became the standard of care, afterwards (Choo, 1995; Torres, 1995; Hammer et al., 1996).

Another major breakthrough was achieved when the first drugs from the new drug classes of PIs and NNRTIs were approved in 1995 and 1996. This allowed for targeting the viral replication cycle and viral genome at different points thereby increasing the genetic barrier once more. This *Highly Active Antiretroviral Therapy (HAART)* containing at least three drugs of at least two classes has been shown to outperform purely NRTI-based combinations in several trials and decreases HIV-related morbidity and mortality dramatically (Hogg et al., 1998b,a; Clavel and Hance, 2004). The use of different drug classes should ensure that the escape mutations of one drug do not provide cross resistance to the other drugs in the regimen which can easily arise when combining for example two NNRTIs. As a consequence, HAART has become standard of care since then. However, despite its success the emergence of resistant viruses is still a major issue. Reasons include the continuous residual viral replication from cells that do not get into contact with the drugs in high enough concentrations. Over time, resistance mutations also accumulate from low-level replication and eventually lead to therapy failure. In addition, adherence of the patients becomes a much more problematic topic in the era of HAART because more drugs in the regimen also mean that patients have to take more pills and have to deal with more side effects (e.g. headache, diarrhea, and lipodystrophy). Thus, it may happen that the patient stops taking some of the drugs or simply forgets to take them. In effect, this lowers the pressure on the virus and allows it to escape to resistance more easily. Adherence problems might also come from different host genetic factors reducing the concentration of the drug in the cell and therefore imposing not enough selective pressure on the virus (Harrigan et al., 2005).

HAART has improved the quality of life of HIV infected people substantially, however HAART is no panacea, and current treatments must be maintained for life (Richman et al., 2009). The long-term side effects of life-long therapy are poorly understood, so far. HIV-associated neurocognitive disorders and HIV-associated dementia are known topics but it is not clear to which extent such effects can be attributed to the virus or the

drugs. Nevertheless, when considering the high amount of pills an HIV-infected patient has to take it is obvious that this cannot be good over time (Wright, 2009; Hawkins, 2010). Therefore, low-toxicity compounds are becoming more and more important and pharmaceutical companies shift their focus more and more towards this field.

## 2.4 Coreceptor Usage and Tropism

### 2.4.1 HIV-1 Tropism and its Discovery

Analysis of HIV-1 coreceptors dates back to the early days of HIV-1 research. After the finding that HIV is the causative agent of AIDS (Gallo and Montagnier, 2003), it soon became clear that the CD4 molecule is a major receptor mediating viral cell entry. However, several observations suggested that in addition to the CD4-receptor, other molecules or receptors must be involved in this process, too (Berger et al., 1998). For example, whereas HIV infection was possible on human cell types that expressed recombinant human CD4 but which would normally not express it, HIV did not infect murine cells with efficient human CD4 expression (Maddon et al., 1986; Alkhatib and Berger, 2007). These results were also found in cell fusion and virus pseudotype assays in which Env was the only HIV component indicating a missing receptor specific to human cells that is needed for HIV cell entry (Ashorn et al., 1990; Clapham et al., 1991).

Another observation was that different viral strains, including strains from individuals at different stages of infection, showed different capabilities to infect either continuous T cell lines or primary macrophages (Fenyö et al., 1994; Miedema et al., 1994; Connor and Ho, 1994). Thus, viruses were designated *T cell line-tropic (T-tropic)* or *macrophage-tropic (M-tropic)*. Viruses capable of infecting both target cell types were referred to as dual-tropic (Berger et al., 1999; Alkhatib and Berger, 2007). This classification also matched very well with another phenotypic difference between viral isolates: the ability to induce syncytia (giant cells that can be seen in light-microscopes) in cultures of peripheral blood mononuclear cells (Asjö et al., 1986). M-tropic isolates usually do not induce syncytia and are therefore also called non-syncytium inducing viruses (NSI) whereas T-tropic viruses are syncytium-inducing (SI), in general (Schuitemaker et al., 1992). Last but not least, it was observed that different viral isolates showed different replicative capacities, in vitro. They could be divided into two major groups, rapid/high and slow/low. While the former can easily be transmitted to a variety of cell lines of T-lymphoid and monocytoid origin, the latter replicate transiently, if at all, in these cell lines (Fenyö et al., 1988). Also this classification seemed to be largely overlapping with the previous phenotypic definitions because slow/low are mostly NSI while rapid/high isolates are in most cases syncytium-inducing (Berger et al., 1998).

All these different classifications for different viruses led to the conclusion that there have to be at least two different *cofactors* or *coreceptors* involved in viral cell entry. The first of these was found in 1996 by a functional cDNA cloning strategy that was used to screen for cofactors that enable murine cells expressing human CD4 to support HIV-1 Env-mediated cell fusion and HIV-1 infection from T-tropic strains. A putative G protein-coupled receptor with seven transmembrane segments, termed *fusin*, was identified. Its importance was further demonstrated by antibodies against it that also blocked cell fusion

and infection in normal CD4-positive human target cells (Feng et al., 1996).

The closest homology of fusin was to a receptor of the subfamily of the CXC-chemokine receptors, a finding which also helped to find the corresponding coreceptor for M-tropic viruses. This was because in 1995, it was shown that RANTES, MIP-1$\alpha$, and MIP-1$\beta$, all members of the chemokine CC subfamily have inhibitory activity against HIV (Cocchi et al., 1995). The inhibitory effect was seen at an early stage of the infection cycle, and mainly against M-tropic but only minimal against T-tropic strains. The conclusion that probably a CC-chemokine receptor with specificity for these chemokines might be the missing cofactor/coreceptor for M-tropic viruses proved to be true a year later, when two independent groups found such a receptor which they termed CC-CKR5 (Combadiere et al., 1996; Samson et al., 1996a). Shortly afterwards, five groups demonstrated that CC-CKR5 is the coreceptor that is needed for cell entry by M-tropic viruses (Deng et al., 1996; Dragic et al., 1996; Alkhatib et al., 1996; Choe et al., 1996; Doranz et al., 1996).

Later on, there was a revision of the chemokine receptor nomenclature leading to CC-CKR5 receiving its current name CCR5, and fusin became CXCR4 (Murphy et al., 2000). The discovery of the chemokine receptors CCR5 and CXCR4 as coreceptors for M-tropic and T-tropic viruses, respectively, also provided an explanation why the observed similarities in the different phenotypic definitions were found before. CCR5 is highly expressed on macrophages which usually do not express CXCR4, thus rendering them specific for M-tropic strains, whereas human continuous T cell lines have a high expression of CXCR4 making them permissive for T-tropic strains (Alkhatib, 2009).

Nowadays, it is known that HIV-1 can use more coreceptors (e.g. Bob, Bonzo, and CCR3) *in vitro*, however so far, only CCR5 and CXCR4 have been shown to be relevant, *in vivo* (Berger et al., 1999; Pastore et al., 2004). Because in the beginning different terms for the biological phenotype existed, these terms are to some extent still used interchangeably. However, almost always people are thinking of HIV-1 coreceptor usage rather than their original meanings when they are using the terms *(biological) phenotype* or *(viral) tropism*.

### 2.4.2 Tropism in the Clinical Context

The classification of HIV-1 into viruses that can use either CCR5 (*R5-viruses*), CXCR4 (*X4-viruses*), or both receptors (*dual-tropic*, sometimes also termed *R5X4-viruses*) is not only an effect that can be seen in *in vitro* experiments but has also several clinical consequences for the patients.

In a newly infected person, usually only R5-viruses can be found, even though the person who infected him might have had X4- or dual-tropic viruses. It is not clear yet, if these viruses are not transmitted or if they cannot establish a new infection because of e.g. the immune system of the newly infected person that recognizes them. Selectivity for R5-viruses was observed independent of the mode of transmission (sexual, blood, or mother-to-child), suggesting that this is not due to different expressions of coreceptors on the target cells (van't Wout et al., 1994; Alkhatib and Berger, 2007).

In accordance with this was also the discovery of a CCR5-allele harboring a 32 base-pair deletion (CCR5-$\Delta$32) within the open reading frame. The allele encodes for a truncated CCR5-protein that is not expressed at the cell surface. Homozygosity for CCR5-$\Delta$32, which can be observed in about 1% of the Caucasian population (Novembre et al., 2005),

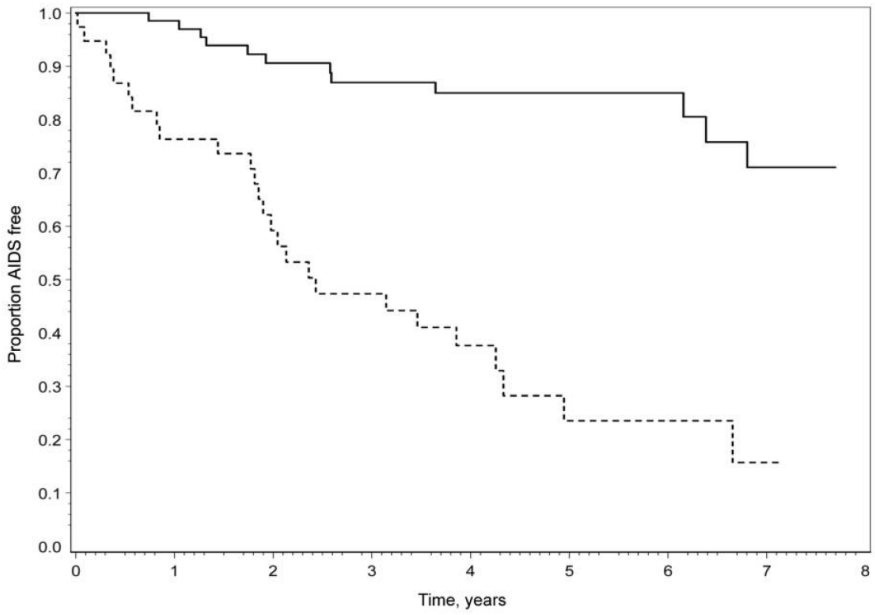


Figure 2.7: Clinical progression to AIDS in patients harboring R5 or X4 viruses. Solid line: R5-viruses, dashed line: X4, dual-tropic, or mixed tropic virus. Figure derived from (Daar et al., 2007).

is associatied with nearly complete resistance to HIV-1 infection whereas heterozygosity is correlated with slower rates of diease progression (Liu et al., 1996; Samson et al., 1996b; Dean et al., 1996).

During the course of infection, CXCR4-using variants are observed in about half of the patients, usually in conjunction with R5-viruses (Melby et al., 2006). The emergence of these viruses is associated with rapid progression to disease (Fenyö et al., 1988; Tersmette et al., 1989; Koot et al., 1993; Richman and Bozzette, 1994; Daar et al., 2007). However, it still remains unclear, if the appearance of CXCR4-using viruses is the cause or the consequence of immune exhaustion (Grivel and Margolis, 1999; Kreisberg et al., 2001). It should also be noted that a switch from CCR5- to CXCR4-usage is predictive but not required for the development of AIDS (Westby and van der Ryst, 2005).

After the initial hunt for the coreceptors and the subsequent association of tropism with different clinical stages, interest shifted towards antiretroviral drugs and the resistance to them. However, the picture changed again in recent years, in which HIV-1 coreceptor usage came into focus once more. This was due to the development of the new class of coreceptor-antagonists. Based on the observation that people homozygous for the CCR5-$\Delta$32 allele have a more or less natural resistance against HIV-1, CCR5 was thought to be an attractive target for drug development against HIV. Naturally, CD4 and CXCR4, being the other (co)receptors used by the virus during cell entry are also interesting targets for new antiretroviral drugs and drugs against them are under investigation. However, CCR5-inhibitors have a major advantage over these: people homozygous for the CCR5-$\Delta$32 deletion have no overt side effects. Thus, there is much more confidence in safety because it seems that in contrast to CD4 and CXCR4, which are crucial for the human, inactivation of CCR5 might be tolerated much better. Not surprisingly, an inhibitor of

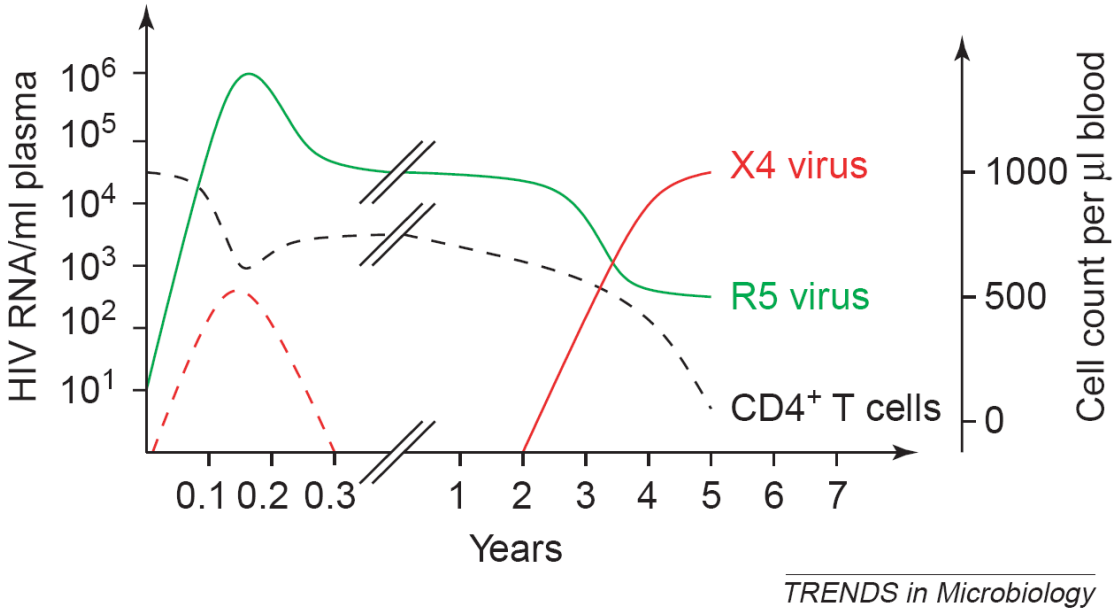


Figure 2.8: Typical progression of R5 and X4 virus variants over the course of HIV infection (from Regoes and Bonhoeffer, 2005). Reprinted with kind permission from Elsevier.

CCR5 (maraviroc) became the first of these to be marketed in 2007 whereas compounds against CD4 or CXCR4 are still far from being licensed.

The excitement about the new class of coreceptor antagonists was mainly based on two reasons. First, coreceptor blocker allow to attack the virus before entering the cell. Coupled with the fact that R5-viruses dominate the early phases of infection, CCR5-antagonists were and still are considered as a possible oral pre-exposure prophylaxis (PrEP) or as vaginal microbicides for women (Dumond et al., 2009; Veazey et al., 2010). The second reason was that for the first time a host rather than a viral protein was the target of an antiretroviral drug. It is commonly believed that this property should hamper development of drug resistance.

In fact, when analyzing data gathered during the drug trials, it became apparent that the development of resistance is rather complicated for the virus as only a small number of patients with resistant viruses have been found. Unfortunately, the virus has a second option to counter CCR5 antagonists: it simply can switch to the CXCR4-receptor. Indeed, about two-thirds of the patients failing in the clinical trials of maraviroc, showed a switch to CXCR4-usage at time of failure (Fätkenheuer et al., 2008; Moore and Kuritzkes, 2009; Cooper et al., 2010).

## 2.5 Tropism Determination

Before administration of an antiretroviral drug, one commonly performs a resistance test in order to determine if the virus is susceptible to the drug or not. In case of coreceptor antagonists resistance plays, as already mentioned before, only a minor role. Instead, since CXCR4-using viruses appear in about 50% of the patients over time, one determines the

tropism of the virus.

Although it might be interesting, at least for certain research questions, to know exactly, if the virus under inspection is R5, X4, or dual-tropic, one usually only differentiates between pure R5- and partly CXCR4-using isolates (sometimes termed *X4-capable*). This is because of several reasons: First of all, the viral population almost always harbors at least some R5-viruses. From a logical point of view this is not surprising, because in the beginning of infection R5-viruses dominate and consequently get archived so that they can appear any time again. Therefore, a CXCR4-antagonist should theoretically also only work in combination with a CCR5-blocker. Another reason is that most assays cannot differentiate between a viral population that contains R5- as well as CXCR4-using viruses, and a population that harbors dual-tropic viruses. Therefore, these populations are also referred to as *dual-mixed* in some phenotypic assays (Coakley et al., 2005). Last but not least, for administration of CCR5-antagonists, it does not matter if only a minor population uses the CXCR4-receptor, and if these viruses are pure X4- or dual-tropic. In any case, the viruses that can use CXCR4 will still be able to replicate in presence of drug pressure and take over the viral population. Hence, in the following we will, if not specifically mentioned, use the term X4 for X4- and dual-tropic viruses as well as for viral isolates in which at least a part of the population can use CXCR4.

In general, two main approaches to assess HIV-1 coreceptor usage exist: phenotypic and genotypic assays. Phenotypic assays measure tropism by growing patient-derived virus *in vitro*. Genotypic assays on the other hand, infer viral tropism from sequence information of the viral envelope protein, usually from the third hypervariable (V3) loop of gp120. Both approaches have their advantages: phenotypic assays are usually more accurate but laborious, expensive and have slow turnaround times, whereas genotypic methods are in general faster, cheaper, and can be performed by many laboratories. The drawback of genotypic prediction of coreceptor usage is that the envelope protein is highly variable making the prediction much more complicated than in the case of resistance testing. With the exception of two well-defined mutations, highly specific for X4-viruses, other mutations seem to have an synergistic effect so that they occur in R5- as well as in X4-viruses.

Over time, a huge array of assays has been developed to determine HIV-1 coreceptor usage. In the following we will only give an introduction to the most important ones used in the clinical context.

### 2.5.1 Phenotypic Approaches

The MT2-assay dates back to the early 1990s when the coreceptors used by HIV-1 were still unknown. It is based on the first *in vitro* coreceptor phenotype assay which classified viruses with respect to their ability to induce syncytia formation in cultures (SI-viruses) of peripheral mononuclear cells (PBMCs). The MT-2 cell line used in the MT2-assay allows for a sensitive, specific and higher-throughput detection of SI-viruses (Koot et al., 1992) than the original PBMC cocultivation assay but it is still very laborious, time consuming and requires highly qualified laboratories (Coakley et al., 2005; Low et al., 2008).

Although the MT-2 assay is a rather old assay compared to the present recombinant virus assays, this does not mean that it performs worse than the latter. In contrast, it has recently been shown to be almost as good as the newest version of the Trofile assay, the

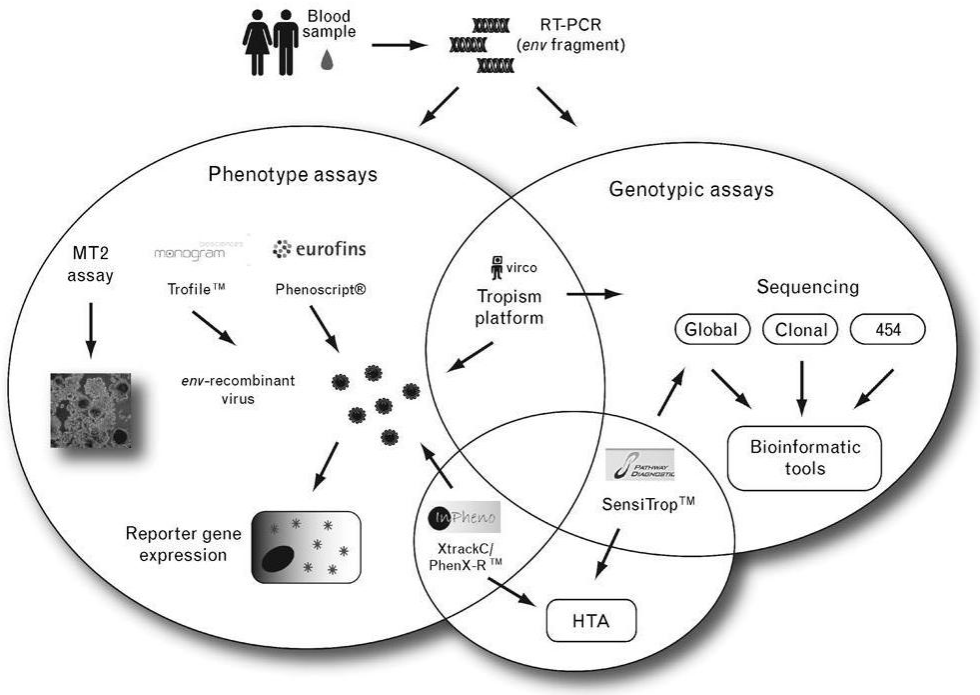


Figure 2.9: Different phenotypic and genotypic approaches to determine HIV-1 coreceptor usage. Derived from (Rose et al., 2009) with kind permission from Wolters Kluwer Health.

most widely used phenotypic test, nowadays (Coakley et al., 2009).

Another type of phenotypic tests for the determination of coreceptor usage is presented by the recombinant phenotypic assays. In these assays, full-length or defined parts of the envelope proteins from a viral sample are amplified and cloned into a plasmid. This plasmid, which also encodes for a reporter gene, is then used to co-transfect cells which in turn then can produce and release pseudoviruses. Dependent on the assays, these viruses are either replication-competent or replication-defective. The pseudoviruses are given to cell lines expressing either of the coreceptors CCR5 or CXCR4. Upon infection of these cells, the reporter gene is expressed and can be measured by light signals or fluorescence (Coakley et al., 2005).

The most prominent representative of these tests is the Monogram Trofile Tropism Assay (*Trofile*) from Monogram Biosciences (South San Francisco, USA) (Whitcomb et al., 2007). It is a single-cycle recombinant virus assay in which full-length patient-derived env gene is amplified by RT-PCR and cloned into an envelope expression vector library that is used to generate luciferase-reporter pseudoviruses. Subsequently these are added to U87 target cells expressing CD4 and either CCR5 or CXCR4. Coreceptor tropism is measured by quantification of emitted light by expression of the luciferase-reporter gene and CCR5- or CXCR4-inhibitors are used as controls (Whitcomb et al., 2007; Coakley et al., 2009).

The Trofile assay has been used in all clinical trials of maraviroc and vicriviroc, and has therefore become the *de facto* gold standard. Recently in June 2008, it has been replaced by a new version, the *Enhanced Sensitivity Trofile Assay (ESTA)* which is able

to reliably detect CXCR4-using viruses down to minorities of 0.3% (Coakley et al., 2009). Disadvantages of the Trofile assay include its long turnaround time of approximately 4 weeks, that it is only available through a central laboratory located in the United States, and the relatively high costs for the assay itself and the logistics linked to it. Furthermore, a proportion of tests fail and therefore produce "non reportable" results, i.e. no result. Last but not least, the viral load of the sample tested must be at least 1000 copies per milliliter. This limitation makes the assays useless for patients which are currently under a successful treatment regimen but which may want to switch to coreceptor antagonists due to e.g. side effects (Swenson et al., 2010b).

In addition to the Trofile assay, a number of other recombinant virus assays have been developed over time. The *Tropism Recombinant Test (TRT) / Phenoscript* (Eurofins VIRalliance, France) uses a methodology similar to the two Trofile assays. The main differences are that a 900bp fragment encompassing the V1-V3 region of gp120 is cloned into the plasmid and that replication competent viruses are produced in HEK293-T cells. Instead of using a luciferase-reporter gene, the pseudoviruses infect indicator cells that carry an HIV-1 LTR-*lacZ* cassette which allows the single-cycle infectivity by a colorimetric assay based on HIV-1 Tat-induced expression of $\beta$-galactosidase (Trouplin et al., 2001; Labrosse et al., 2003). The advantages and disadvantages of the assay are quite similar to the ones of the Trofile assay.

### 2.5.2 Genotypic Approaches

Genotypic approaches comprise two steps: the sequencing method and the prediction method interpreting the sequence data. Sequencing is normally based on standard population based ("bulk") Sanger sequencing. In this case, a consensus sequence of the viral quasispecies is determined. The sequence can contain nucleotide mixtures (positions at which different nucleotides are present in the quasispecies) complicating the interpretation. Furthermore, only mutations being present in at least 15-20% of the viral population can reliably be detected with bulk sequencing methods.

Recently, a new method termed massively parallel sequencing became available. With this method, one can sequence hundreds of thousands of viruses in parallel. Instead of a "bulk"-genotype, the output of this method is a set of individual sequences, each representing one clone in the viral population. Thus, the technology allows for looking "deeply" into the quasispecies (therefore it is also referred to as *(ultra)deep sequencing*) and so a much more fine-granulated picture can be determined. Not only are single clones determined rather than a consensus sequence but also can viruses prevalent in minor populations reliably be sequenced with this new technology. We will introduce massively parallel sequencing in more detail in Chapter 4.

Interpretation of the genotypes is performed either with simple rules or with more sophisticated machine learning methods. All of them are based on sequence information of the V3 loop which is known to be the major determinant of coreceptor usage (Hwang et al., 1991).

One of the simplest but still a very popular method of predicting coreceptor usage is the so-called *11/25-rule*. It predicts a virus to be X4, if a positively charged residue is found at position 11 and/or 25 of the V3 loop (Fouchier et al., 1992). The 11/25-rule is quite

accurate for R5-viruses but misclassifies many X4-viruses (Jensen et al., 2003).

The charge of the loop was found to be predictive, in general. Briggs et al. generated a linear regression model based on the number of positively and negatively charged residues, the overall charge of the loop, and the presence of an isoleucine residue at position 292 of the envelope protein gp120 (Briggs et al., 2000).

Further progress was made by using several different statistical learning methods. The first models used decision tree classifiers (Beerenwinkel et al., 2001; Pillai et al., 2003) or artificial neural networks (Resch et al., 2001), later on Support Vector Machines (SVMs) (Pillai et al., 2003) and Position Specific Scoring Matrices (PSSMs) were introduced into the field (Jensen et al., 2003, 2006). More recently, people tried to use Mixtures of Localized Rules (Sing et al., 2004) and logistic regression models with additional sequence features (Prosperi et al., 2009).

In principle, all of these models rely on a correct alignment of the viral genotype against a reference or a consensus sequence. Because the V3 loop is highly variable and shows insertions and deletions frequently, this can lead to problems. Consequently, one group developed a prediction model using *String Kernels* which do not suffer from these deficiencies (Boisvert et al., 2008).

Last but not least, additional information can be incorporated into these prediction systems. Sander et al. have shown how to incorporate properties of the three-dimensional structure of the V3 loop. They improved over prediction methods working on sequence information alone by encoding structural information into a descriptor that lists distance distributions between functional groups in the loop (Sander et al., 2007). Another recent structure based approach has been described by Dybowski et al. (Dybowski et al., 2010). In another work, clinical markers such as viral load and $CD4^+$ T cell counts were added to the statistical model (Sing et al., 2007).

Despite the larger number of methods developed for predicting HIV-1 coreceptor usage, only a small number of web-services exist. The most commonly used are WetCat, WebPSSM, and geno2pheno[coreceptor]. WetCat implements the charge rule, three different decision trees, and Support Vector Machines (Pillai et al., 2003). WebPSSM offers different Position Specific Scoring Matrices for working with subtype B and C samples. geno2pheno[coreceptor] also uses SVMs to infer coreceptor usage but in contrast to the other methods, can deal with sequence ambiguities and with nucleotide as well as amino acid sequences. In addition, the latter method allows for changing the specificity-level of the prediction method and is the only system that can deal with clinically derived markers such as $CD4^+$ T cell counts.

### 2.5.3 Other Approaches

Some methodologies to determine coreceptor usage are neither clearly phenotypic nor entirely genotypic. For example, Virco BVBA is developing a platform combining both approaches. It consists of four assays in total: population sequencing of the V3 loop, clonal sequencing of the $NH_2V4$-fragment (the N-terminal fragment of gp120 to V4), and population and clonal phenotyping using $NH_2V4$-recombinant viruses (Baelen et al., 2007).

A different approach to assess viral tropism is heteroduplex tracking analysis (HTA) which is used in the SensiTrop QT assay (Pathway Diagnostics, Malibu, US). In this case,

X4 viruses are determined by the formation and separation of heteroduplexes to CCR5 V3 molecular probes. R5-results by HTA are further investigated by sequence analysis. The assay has a very short turnaround-time of below 7 days (Rose et al., 2009). Related to the approach used in the SensiTrop assay is the Xtrack$^C$/PhenX-R assay by the InPheno AG (Basel, Switzerland). This assay also consists of two methodologies: a probe-based hybridization assay (Xtrack$^C$) and a phenotyping assay (PhenX-R). The former is used to rapidly test the sample with R5-specific and X4-specific fluorescence-labeled probes, with distinct capillary migrations. In the case of ambiguous samples or a suspected mix of tropism, the phenotypic assay which is similar to the Phenoscript assay, is performed afterwards. The main advantage of this approach is the ability to screen most samples with the rapid and inexpensive Xtrack$^C$ and relying on the phenotypic test only in complicated cases (Rose et al., 2009).

### 2.5.4 Performance Comparison Between Different Methods

It is relatively hard to assess the true quality of the different methods determining coreceptor usage. In general, these assays are validated on a set of well-defined viral strains which do not necessarily reflect the "real world" clinical samples harboring a whole population of different viruses. Also mixing experiments in which two viral strains are mixed together at defined quantities in order to find the detection limit of the assays should be considered carefully because often very well-defined strains are used which are much easier to detect than most viruses that occur in the routine setting. To account for this, one commonly tries to compare the assays against a *gold standard*. Because there is no perfect gold standard, often the Monogram Trofile Tropism Assay is used as reference because it has been used in all clinical trials to date. Naturally, something people often forget, the Trofile assay is also not guaranteed to be always correct. Otherwise, the assay would not have been replaced by a "better" enhanced version which is nowadays considered to be the gold standard.

Comparisons of different methods have mainly been performed between genotypic prediction methods, on the one hand, and the Trofile assay, on the other. However, also some comparisons between different phenotypic assays are available. For example in 2007, the Trofile assay was compared against the TRT and an older version of the current geno2pheno[coreceptor] system. On a set of 74 antiretroviral-naïve patients, the concordance with the two phenotypic assays was 85.1%, without evidence of a difference in sensitivity. The concordances to the bioinformatics method were 86.5% (Trofile) and 79.7% (TRT), respectively, approaching the degree of agreement between the two phenotype assays (Skrabal et al., 2007). A more recently developed assay, the *Toulouse Tropism Test (TTT)* achieved a concordance of 91.7% with the Trofile assay on a dataset of around 600 samples (Raymond et al., 2010). The same group also offered an explanation for the differences observed between different assays: there could be sample bias introduced into the virus population during PCR amplification that leads to different populations being genotyped or phenotyped in the respective assays. Using the same bulk env PCR products to perform genotypic and phenotypic assays in parallel, Raymond et al. found much higher sensitivities (88%) at high specificities (87%) for geno2pheno[coreceptor] than seen in other works (Raymond et al., 2008).

The SensiTrop assay has been shown to detect and quantify X4 viruses down to levels as low as 1% in mixing experiments. However, in a blinded assessment comparing 100 samples with the Trofile assay, SensiTrop detected an CXCR4-using virus in less than 50% of the samples that were phenotyped as CXCR4-capable by the Trofile assay (Low et al., 2008). An improved version of the assay was recently reported to have an overall concordance of 79% (Swenson et al., 2010a).

Performance comparisons of different genotypic methods have been reported in several works, lately. These comparisons can be divided into two groups: performance on *clonal* and *clinical* samples. In clonal samples, there should be only a single viral strain in the isolate so that there can be neither a problem with minority detection by the sequencing method nor can there be a sampling problem. Instead there should be a single clear genotype which matches to the phenotype. In case of clinical samples, we refer to a normal patient sample containing a whole viral population. Coreceptor usage of these samples is much more complicated to determine both for genotypic as for phenotypic methods because they can contain CXCR4-using variants at very low numbers and there might be a sampling bias when comparing different methods. Thus, it can happen that one method still "sees" the X4-viruses while the other does not. Standard population-sequencing is thought to detect minorities only as low as 15-20% whereas phenotypic methods can reliably detect minorities of 5% or even less. Therefore, even a perfect bioinformatic interpretation of the genotype might produce "wrong" results in clinical samples but should show very good results on clonal samples in which these problems do not occur.

In fact, when tested on clonal samples, genotypic methods showed much better results than on clinical samples. When tested on data from the Los Alamos HIV Sequence Database that contains mainly clonal isolates, the 11/25-rule showed a sensitivity of 59.5% in detecting X4-viruses at 92.5% specificity. At the same specificity, PSSMs and Support Vector Machines showed sensitivities of 71.9% and 76.9%, respectively (Sing et al., 2007). Prediction models incorporating structural information were evaluated on a subset of the Los Alamos dataset containing only sequences without deletions and insertions. In comparison with models using sequence information alone, they increased the sensitivity to detect X4-viruses by 7 percentage points (Sander et al., 2007).

In contrast to these results, initial results on clinically derived isolates were very discouraging. Low et al. compared different prediction systems of coreceptor usage on 977 antiretroviral naïve samples of the British Columbia HAART Observational Medical Evaluation and Research (HOMER) cohort. The sensitivities of different methods in comparison with the Trofile assay ranged from 30% to 50% at specificities of around 90%. This led to the conclusion that genotypic methods are inadequate for clinical use (Low et al., 2007). When testing prediction models utilizing clinical correlates like $CD4^+$ T cell counts, much better results where achieved on the same dataset. At a specificity of 93.5%, these methods reached sensitivities of up to 63% corresponding to a 2.4-fold improvement in detecting X4-viruses relative to the 11/25-rule (Sing et al., 2007).

# 3 The World World Beyond V3 - Extending the Region of Interest

The envelope gene of HIV encodes for the glycoproteins gp120 and gp41. Both proteins can be further divided into regions (see Figure 3.1). In gp120, these are the five variable (V1-V5) and five conserved (C1-C5) regions. While HIV-1 is known for its high mutation rate in general, this is even more the case in the envelope, especially in its variable regions. These encode for variable loops that allow not only for several substitutions but also for multiple insertions and deletions. In contrast to them, the conserved regions are relatively stable but can still mutate substantially. While gp120 is subdivided into regions based on their variability, gp41 is subdivided into functional regions (see Figure 3.1).

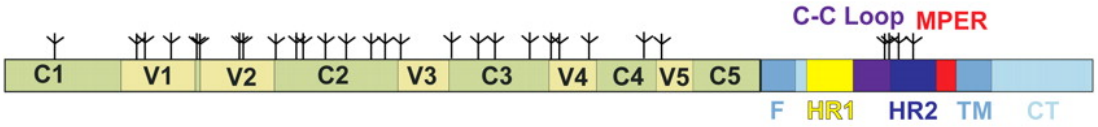


Figure 3.1: Schematic representation of the HIV-1 envelope gene. The conserved regions C1-C5 and the variable regions V1-V5 belong to gp120 whereas *fusion peptide* (F), *heptad repeat 1* (HR1), *immunodominant loop with a conserved disulfide bond* (C-C loop), *heptad repeat 2* (HR2), *transmembrane anchor* (TM) and *cytoplasmic tail* (CT) are functional regions of the gp41 protein. Figure derived from (Frey et al., 2008). Copyright (2008) National Academy of Sciences, U.S.A.

As already mentioned in the previous chapter, the third hypervariable loop (V3 loop) of the envelope protein gp120 is known to be the major determinant of coreceptor usage. The correctness of this principle is reflected by the fact that coreceptor usage prediction based on the V3 loop works very well when tested on matched genotype-phenotype pairs derived from clonal isolates (Sing et al., 2007). However, in the course of the years, several mutations outside V3 have been shown to be correlated with HIV-1 coreceptor tropism, too. While some of these mutations might arise from covariation with mutations within V3, e.g. to relieve sterical tensions or to re-establish fitness losses, other mutations may have an effect on tropism directly. Some cases are known where the same V3 loop was observed in samples of different phenotypes. This can only be explained by tropism-defining mutations located outside of the V3 loop.

Within this chapter, we will try to explore the use of different regions within the envelope proteins gp120 and gp41 and even go further by analyzing also other proteins of HIV-1 and their impact on coreceptor usage. A major problem for this analysis is that we only have limited data from regions beyond V3. The main reason for this is that the V3 loop alone

already explains most of the phenotypes and the effects we see in the patients. Hence, it is generally regarded as inefficient and too expensive to sequence other regions of the viral genome as well. As a consequence, we have only a small number of samples (if any) with genotypic and phenotypic information for other regions. Another point limiting the size of the datasets is that some regions are very hard to sequence because of the high variability and in some parts extensive length polymorphism of the envelope protein. For clonal data this might be a minor problem but clinical isolates containing strains of different lengths can be very complicated to interpret (see e.g. Figure 3.2). Of course, a prediction system making use of several parts or the whole envelope protein would be desirable but unfortunately in most cases only smaller parts rather than the whole protein are sequenced so that we do not have enough data to generate a reliable model from these.

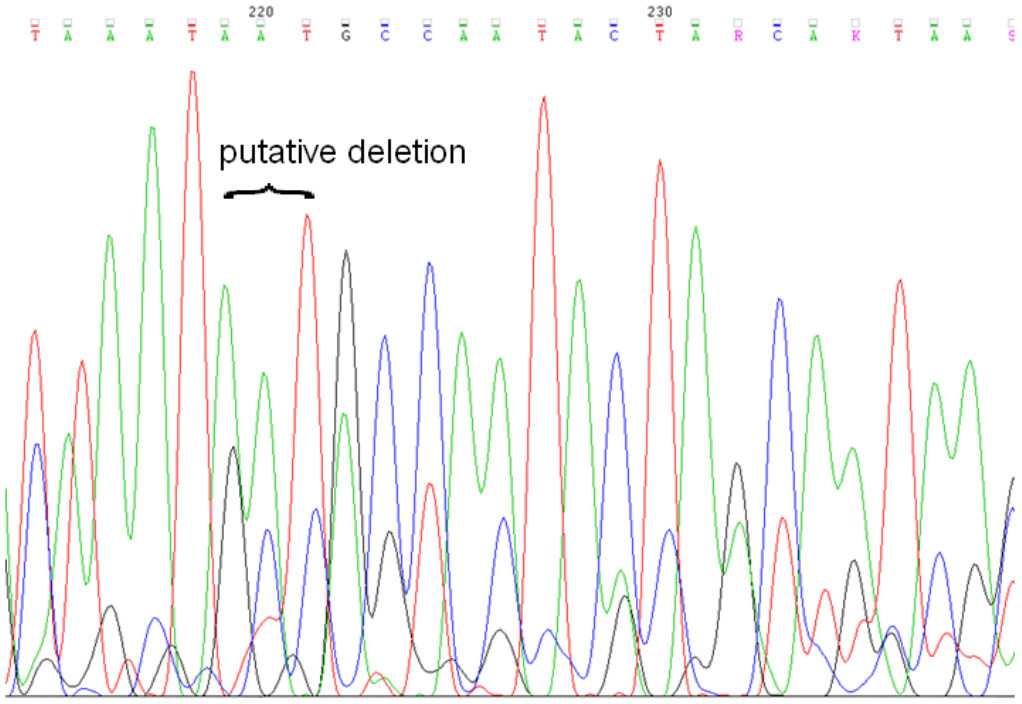


Figure 3.2: Typical electropherogram of the V2 loop. Viral strains of different lengths lead to profiles that are hard to interpret. In this case the minor variant (the one with the second highest peaks) has a deletion of just a single amino acid compared to the major strain. Although the two variants differ minimally aside from this, the overall electropherogram looks very messy.

The order in which we will discuss the different regions does not reflect their importance nor their order in the envelope gene, but is based on data availability and the time at which large enough datasets with clinical data became available to us.

## 3.1 Improving Predictions with the Second Hypervariable Loop of gp120

HIV-1 coreceptor usage is mainly determined by the V3 loop and prediction of coreceptor usage from the V3 loop has been shown to be a valuable and reliable resource. However, the fact that viral strains with identical V3 loops but different tropism have been observed suggests that mutations outside V3 affecting coreceptor usage must exist. In one of the first works showing this to be true, Andeweg et al. transferred fragments of SI viruses to NSI envelope genes. They found that not only fragments covering the V3 region of SI viruses but also some fragments not including the V3 loop could transfer syncytium-inducing capacity to envelope proteins previously not capable of inducing syncytia (Andeweg et al., 1993).

Several other reports confirmed these results and could determine specific amino acid substitutions rather than regions influencing HIV-1 coreceptor usage (e.g. Groenink et al., 1993; Hoffman et al., 2002; Boyd et al., 1993; Huang et al., 2008). Most determinants affecting viral tropism outside V3 were found in the bridging sheet, a structural region formed by the variable loops 1 and 2 (V1 and V2) and the fourth conserved region (C4). This finding is not especially surprisingly because the bridging sheet covers substantial parts of the coreceptor binding site (Kwong et al., 1998; Biscone et al., 2006). Among the regions of the bridging sheet, the second hypervariable (V2) loop of gp120 has been reported most frequently and seems to be the most important.

The V2 loop can modulate coreceptor usage both through an interaction with the V3 loop leading to an alteration in the whole conformation of gp120 (Koito et al., 1994), and via V3-independent mechanisms (Koito et al., 1995). Some authors also suggested that mutations within V2 may compensate for loss-of-fitness mutations in the V3 loop that are necessary but not sufficient for coreceptor switching (Pastore et al., 2006). Altogether, the V2 loop seems to be a promising region for incorporation in prediction systems of HIV-1 coreceptor usage.

Although mutations in the V2 loop have been shown to influence HIV-1 coreceptor usage in several experiments, this was to our knowledge neither shown on larger datasets nor used to improve coreceptor usage prediction. Hoffman et al. developed an elegant approach to analyze a larger dataset despite having only a limited amount of data. They used data from the Los Alamos Sequence Database irrespective of available phenotypes. Instead of an experimentally determined phenotype, they classified the sequences into R5-like and X4-like, based on an adapted 11/25-rule (an arginine at position 25 in V3 was not used to discriminate the phenotypes). Using these inferred phenotypes, they could confirm previous observations and predict specific positions contributing to viral tropism (Hoffman et al., 2002). For the development of a coreceptor usage prediction method incorporating sequence data from other regions than V3, however, this approach might not be perfect because of two reasons. First, the inferred tropisms might be in fact incorrect and drive the prediction system into the wrong direction. With respect to wrongly classified R5-viruses the risk is however limited because the 11/25-rule is known to be highly specific and its specificity has even been increased with the approach of Hoffman et al. due to the removal of the arginine at position 25. The second, more problematic point is that the identified mutations are not independent of known V3-mutations but correlated. Because they appear together with

the 11/25-mutations, we cannot expect any significant impact from them as we already can infer tropism from these sequences with the simple 11/25-rule. Instead, our interest lies more in mutation pathways towards CXCR4-usage which evolve in the absence of known V3-mutations. Based on these considerations, we refrained from the described approach in this section and based our analysis on datasets of experimentally determined phenotypes. The aim was to find mutations and additional properties significantly associated with coreceptor usage, and to use these to improve prediction systems, both for clonal as well as for clinically derived isolates.

The work described in this section was carried out in collaboration with the Institute of Virology at the University of Cologne (Cologne, Germany) and the British Columbia Center for Excellence in HIV (Vancouver, British Columbia, Canada). It has been published in *Journal of Infectious Diseases* under the title "Improved prediction of HIV-1 coreceptor usage with sequence information from the second hypervariable loop of gp120" (Thielen et al., 2010b).

### 3.1.1 Material & Methods

**Datasets** The analyses in this study are based on three distinct datasets of different characteristics. The first dataset, termed *clonal* included all samples of the Los Alamos HIV Sequence Database (http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html), July 2007) having sequence information from the V2 and the V3 loop as well as an experimentally determined phenotype. The dataset consists of 916 sequences from 312 different patients. We refer to the dataset as "clonal" since the majority of the data is derived from clones and contains no associated clinical data. Strictly speaking, not all of the dataset is purely clonal. Upon checking the annotations as far as possible, more than 85% of the samples were indeed of clonal origin, while about 10% seem not to be molecular or biological clones and for around 5% of the samples we could not find the respective information. To ensure that this does not induce a bias into our analysis, we compared the results of the mutation analysis with a similar one based only on the samples we found to be clonal in their annotations. In this analysis, only marginal differences were observed and correlation of the frequencies of the significant mutations was 0.9955 between the two datasets.

The second dataset, termed *therapy-naïve* contained 268 therapy-naïve samples from a subset of the well described British Columbia HAART Observational Medical Evaluation and Research (HOMER) cohort (Hogg et al., 2001). This cohort consists of HIV-positive, antiretroviral-naïve adults who started triple antiretroviral therapy through the B.C. Drug Treatment Program between August 1996 and September 1999. The samples used in this study were the latest ones collected in the 180 days prior to therapy initiation (Hogg et al., 2001). Coreceptor usage was determined by the original Trofile assay. The V2 and V3 loops were "bulk"-sequenced from the same baseline plasma samples as described previously (Brumme et al., 2005; Low et al., 2007).

The third (*therapy-experienced*) dataset consisted of 64 therapy-experienced patient isolates collected at the Institute of Virology of the University of Cologne. These samples were screened for maraviroc administration in routine diagnostics. Genotypes of the V2 and V3 loops were "bulk"-sequenced methods and the phenotype determined with the original Trofile assay (Whitcomb et al., 2007).

**Determination of envelope sequences of clinically derived isolates** Genotypes of the V2 and V3 loops were determined by population sequencing. Isolates from the patients in the therapy-naïve dataset were sequenced as described in (Brumme et al., 2005). Viral RNA of samples collected for the therapy-experienced dataset was isolated from plasma by using the MagNA Pure Compact Nucleic Acid Isolation Kit (Roche Applied Science, Mannheim, Germany) according to manufacturer's protocol. Reverse transcription and PCR were performed using the OneStep RT-PCR kit (Qiagen, Hilden, Germany) with the primers V3F: 5'-AGAGCAGAATTCAGTGGCAATGAGAGTGA-3' (nt 6202-6222) and V3R: 5'-AGTGCTTCCTGCTGCTCCYAAGAACCC-3' (nt 7785-7759), followed by a nested PCR with HotStartTaq (Qiagen, Hilden, Germany) with primers V3-1: 5'- TGGGATCAAAGC-CTAAAGCCATGTG -3' (nt 6558-6582) and V3-5: 5'- AAAATTCCCCTCCACAATTA -3' (nt 7371-7352). PCR-products were purified using QIAquick spin PCR purification kit. The sequence reaction was performed with the primers V3-1, V3-2: 5'- CAGTACAATGY-ACACATGG -3' (nt 6955-6973), V3-3: 5'- CCATGTGTRCATTGTACTG -3' (nt 6955-6937), V3-4: 5'- GTACAATGTACACATGGAAT -3' (nt 6957-6976), V3-5 and the ABI reaction mix. Extension products were purified using MultiSreen purification plates (Millipore, Bedford, MA, USA) and Sephadex G-50 superfine (Amersham Biosciences, Uppsala, Sweden) and were run on an ABI 3130 Avant capillary sequencer. The obtained sequences were assembled and edited using the Seqman software implemented in the DNASTAR Lasergene suite (DNASTAR Inc., Madison, WI, USA).

**Sequence alignment and encoding** V2 and V3 loop sequences of the clonal dataset were aligned with ClustalW (Thompson et al., 1994) using standard parameters following manual inspection. Due to the extensive length polymorphism in V2, we inserted several gap columns in the V2-alignment before positions 163 and 187 (relative to HXB2, see Figure 3.3). The respective alignments were then used as sequence profiles to which the samples of all three datasets were sequentially aligned.

A canonical indicator representation was used to encode the aligned sequences. Each position in the aligned sequence was defined by a binary vector of length 21, in which each component indicated the presence or absence of a specific amino acid or a gap at this specific position.

**Data analysis, model generation and evaluation** We trained Support Vector Machines (SVMs) for coreceptor usage prediction using the SVM$^{light}$-package (Joachims, 1998). SVMs are supervised learning methods used for classification and regression. They map the input data into a high-dimensional feature space in which it can be optimally separated by a hyperplane into two categories, in this case R5- and X4-viruses. The hyperplane is then used to classify unseen data. The mapping is done with a kernel function that can be linear and non-linear (e.g. a polynomial or a radial basis function). Because previous studies showed that non-linear kernels do not improve prediction quality for simple encodings (Sing et al., 2004) we used linear kernels in this work, as well. Using an internal grid-search, we optimized the cost factor $c$ within a range from $c = 10^{-6}$ to 10.

The difference between the number of samples and the number of patients could lead to a bias due to overrepresented patients with many (similar) sequences. We tried to minimize these effects by repeating all of our analyses on the clonal dataset ten times. In every

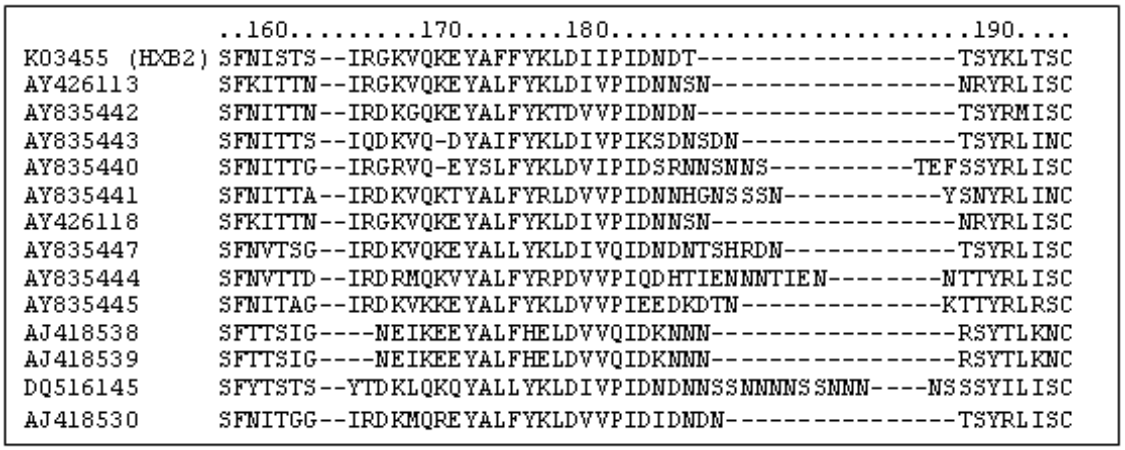


Figure 3.3: Excerpt of the V2 sequence profile. Because of the extensive length polymorphism, gap columns have been inserted after HXB2-position 186.

replicate at most one R5- and one X4-sequence per patient was randomly chosen.

Performances of different prediction models on the clonal dataset were assessed using 10-fold cross validation. Furthermore, models trained on the clonal dataset were evaluated on the datasets containing clinically derived samples. To afford direct comparisons, models containing the V2 loop were trained and evaluated on the same set of samples as models only using V3 in all prediction comparisons.

In addition to the prediction models trained in this section, we also compared them with the 11/25-rule (Jensen et al., 2003), and the two web-systems geno2pheno[coreceptor] (Lengauer et al., 2007) and WebPSSM (Jensen et al., 2003). Statistical analysis was performed using the statistical programming language R (Team, 2005) and the R-package ROCR (Sing et al., 2004). The area under the ROC-curve (AUC), sensitivity, and specificity were used as performance measurements. Significance of different AUCs was assessed using the approach described by DeLong et al. (DeLong et al., 1988).

**Sequence analysis** Mutations significantly associated with HIV-1 coreceptor usage were determined using sequences from the clonal dataset. We tested every amino acid at every position within V2 and V3 for correlation with viral tropism. Because of the extensive length polymorphism in V2, the positions from 31 to 51 (residues between position 185 and 189 in HXB2) in the sequence profile of V2 could not be aligned very well (see Figure 3.3). We did not find any clear sequence pattern that could serve as a scaffold for the alignment. Thus, the positions of the aligned residues were quite uncertain and a mutation analysis based on fixed positions would not have made sense for these. We therefore excluded this *length-polymorphic* part of V2 (in contrast we call the remaining positions the *V2-stem*) from the mutation analysis. Significance of amino acids at specific positions was assessed with Fisher's exact test. Only mutations that were significant ($p < 0.05$) in all ten replicates, were taken as predictive in the results.

We also tested further features for their correlation with R5- and X4-viruses. These were the number of positively (negatively) charged residues, the overall charge of the regions, the number of N-glycosylation sites, and the length of the respective region. Because these

features are position-independent they do not rely on a correct alignment. Therefore, in contrast to our mutation analysis, we could use all residues of the V2 loop in this analysis. In addition, we did not only assess the predictivness of these features for the complete V2 and V3 loops, but also for the conserved and polymorphic parts of V2 alone. Significance of position-independent features was determined using Student's t-test. Features that are reported to be significant had to be significant ($p < 0.05$) over all ten runs of the analysis.

### 3.1.2 Results

#### Clonal dataset

**Impact of specific mutations on coreceptor usage** We found several amino acids within V2 and V3 to be significantly correlated with one of the two phenotypes in the clonal dataset. The results are shown in Table 3.1 and Appendix Table 1. The displayed residues were significant in each of the ten replicates containing at most one R5- and one X4-sequence per patient (see Materials and Methods - Data analysis, model generation and evaluation). Frequencies, log odds ratios, and the 95% confidence intervals were computed from the whole dataset.

| *POS* | *HXB2* | *MUT* | *fR5 (%)* | *fX4 (%)* | *LOR* | *95% CI* |
|---|---|---|---|---|---|---|
| 161 | I | A | 8.05 | 1.1 | -2.94 | [0.037, 0.074] |
| 164 | S | V | 4.24 | 10.99 | 0 | [0.704, 1.419] |
| | E | E | 55.08 | 30.77 | -1.53 | [0.145,0 .319] |
| 165 | I | L | 40.25 | 25.27 | -1.41 | [0.166, 0.353] |
| 166 | R | I | 1.27 | 7.69 | 0.84 | [1.648, 3.302] |
| 169 | V | T | 0.42 | 13.19 | 2.48 | [8.444, 17.05] |
| | K | K | 45.34 | 25.27 | -1.53 | [0.146, 0.314] |
| 173 | Y | Y | 72.46 | 54.95 | -1.22 | [0.184, 0.463] |
| | H | H | 14.41 | 27.47 | -0.3 | [0.508, 1.062] |
| 177 | Y | Y | 96.61 | 89.01 | -1.03 | [0.135, 0.928] |
| 182 | I | E | 1.27 | 7.69 | 0.84 | [1.648, 3.302] |
| 184 | I | L | 24.15 | 13.19 | -1.55 | [0.147, 0.301] |
| 192 | K | R | 87.29 | 68.13 | -1.2 | [0.171, 0.529] |
| | I | I | 5.93 | 15.38 | 0 | [0.701, 1.425] |
| 195 | S | N | 69.07 | 45.05 | -1.38 | [0.163, 0.388] |
| | H | H | 2.54 | 15.38 | 0.84 | [1.638, 3.323] |

Table 3.1: Significant mutations within V2. POS: position in HXB2, HXB2: residue at this position in HXB2, MUT: residue tested for association with phenotype, fR5 & fX4: frequencies in R5- and X4-viruses, respectively, LOR: Log odds ratio, 95% CI: 95% confidence intervals

It is obvious that the frequencies of significant residues within the V2 loop of R5- and X4-isolates did not differ as much as in the third hypervariable loop of gp120 (see also Sing et al., 2007). Instead, the respective amino acids within V2 occurred in R5- as well as in X4-isolates in most cases and only trends between the two groups but not large differences

could be seen for individual mutations. Nevertheless, there were also mutations which were relatively specific for one of the two phenotypes. For example, the mutations V169T or S195H seemed to be highly predictive for viruses using the CXCR4-coreceptor.

**Impact of position-independent features on coreceptor usage** We found several position-independent features to be significantly different between the two phenotypes (see Table 3.2). In agreement with our expectations, we found the number of positively and negatively charged residues, the number of N-glycosylation sites, and the charge of the V3 loop to be significantly different between R5- and X4-viruses. The number of negatively charged residues as well as the overall charge also differed significantly when analyzing sequence data from the second hypervariable loop of gp120. However, despite showing a broad variation (see Figure 3.4), the length of the V2 loop did not prove to be correlated with R5- or X4-samples.

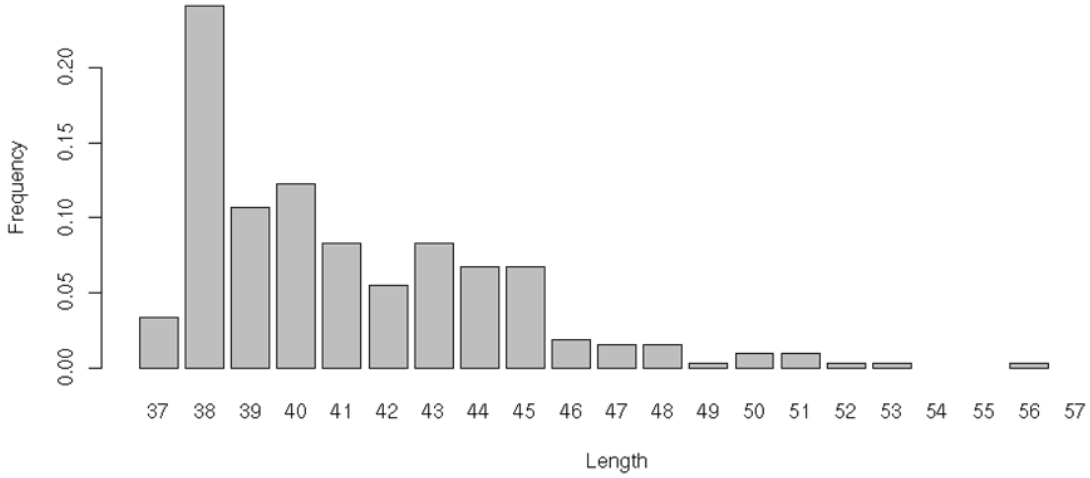


Figure 3.4: Extensive length variation of the V2 loop. The distribution of different V2 loop sizes is shown. About 25% had an average length of 38 residues but also sequences with up to 56 amino acids were found in the dataset.

When focusing on the subregions of V2, we were surprised to see that these differences did not appear in the conserved part but that only the number of N-glycosylation sites differed there significantly. Because we could not find any clear sequence pattern in the polymorphic subregion, we also did not expect to find any other significant result in this region. However, in contrast to our expectations, all other tested features were significantly different between R5- and X4-samples in the polymorphic region: the number of positively and negatively charged residues, the overall charge and the length of the region.

**Comparison of Prediction Methods** The differences within V2 observed with respect to amino acids and position-independent features suggested incorporating the second hypervariable loop of gp120 into coreceptor prediction methods. Therefore, we generated three different prediction methods using sequence information either from the V2 loop alone, from the V3 loop alone, or from both loops together. Validation was performed with ten

| ***region*** | ***feature*** | ***∅ in R5*** | ***∅ in X4*** | ***p-value*** |
|---|---|---|---|---|
| V3 | # pos. charged amino acids | 6.08 | 7.86 | $< 0.0001$ |
| | # neg. charged amino acids | 1.69 | 1.24 | $< 0.0001$ |
| | charge | 4.39 | 6.62 | $< 0.0001$ |
| | # N-glycosylation sites | 0.98 | 0.59 | $< 0.0001$ |
| | length of region | 34.90 | 35.08 | 0.1532 |
| V2 | # pos. charged amino acids | 6.37 | 6.75 | 0.0718 |
| | # neg. charged amino acids | 4.95 | 4.43 | 0.0004 |
| | charge | 1.41 | 2.31 | 0.0001 |
| | # N-glycosylation sites | 1.61 | 1.65 | 0.6959 |
| | length of region | 41.58 | 42.38 | 0.0747 |
| V2-stem | # pos. charged amino acids | 5.73 | 5.72 | 0.9381 |
| | # neg. charged amino acids | 3.10 | 2.95 | 0.1201 |
| | charge | 2.63 | 2.77 | 0.3926 |
| | # N-glycosylation sites | 0.91 | 0.81 | 0.0247 |
| | length of region | 32.95 | 32.86 | 0.2301 |
| V2-polymorphic | # pos. charged amino acids | 0.63 | 1.03 | 0.0037 |
| | # neg. charged amino acids | 1.85 | 1.48 | 0.0006 |
| | charge | -1.21 | -0.45 | $< 0.0001$ |
| | # N-glycosylation sites | 0.69 | 0.83 | 0.1981 |
| | length of region | 8.63 | 9.51 | 0.0495 |

Table 3.2: Differences in properties among R5- and X4-viruses in different regions of the gp120 envelope protein.

replicates of ten-fold cross validation. In each cross-validation run, the pool of training and test samples was the same for all three predictors.

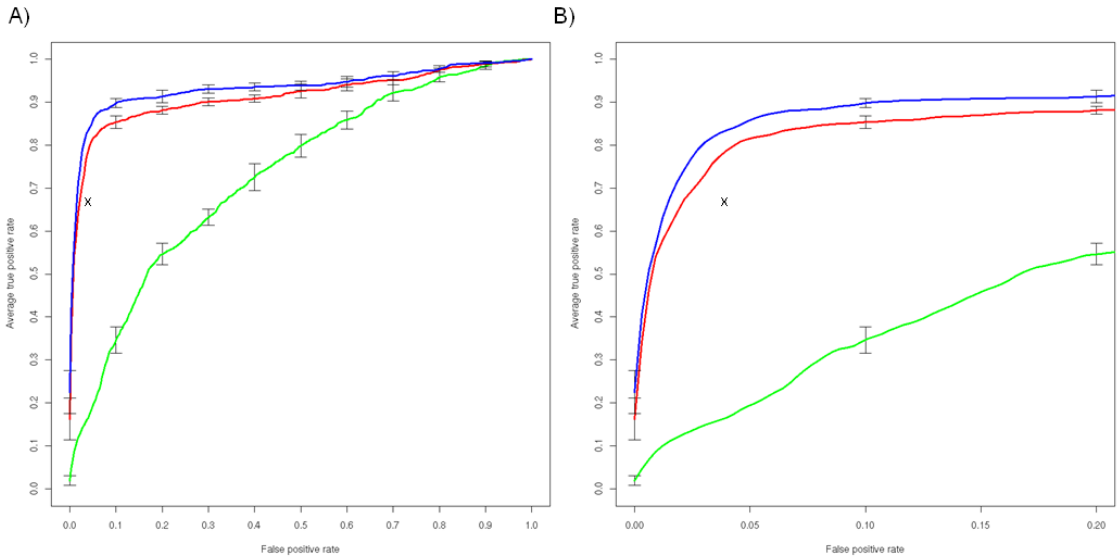


Figure 3.5: Comparison of prediction results on the clonal dataset using either sequence information of the V2, the V3, or both loops together. The blue line characterizes the ROC-curve of the predictor using the V2 and V3 loop, the red line the one of the predictor using V3 only, and the green line the prediction results from the predictor using sequence information of the V2 loop alone. The black cross marks the performance of the 11/25-rule. Variation around the average curve is visualized with standard error bars. A) shows the overall results while B) focuses on the sensitivities at high specificities. The curves from the ten replications were averaged using threshold averaging.

The ROC-curves of the three different prediction methods are depicted in Figure 3.5. Models that were trained only on sequence information from the V2 loop were clearly outperformed by the other two models and the 11/25-rule. On average, the V2-models could only detect 34.6% of the X4-isolates with a specificity of 90% (i.e. when allowing 10% of the R5-samples to be falsely predicted as X4). 90% specificity is usually used for comparisons of coreceptor usage predictions. In comparison to the 11/25-rule this is very low since the rule correctly predicted 66.7% of the X4-isolates by having a higher specificity of 96.1% (see the x in the figure). Nevertheless, the average area under the ROC-curve of the V2-predictor was with 0.76 well above 0.5 which would correspond to a random predictor. Thus, a possible positive influence of the V2 loop could be expected when using it in combination with the V3 loop.

Prediction models trained on V3 sequence data alone showed much better results than the V2-models and the 11/25 rule. On average, sensitivity of detecting X4-viruses was 85.3% at a specificity of 90%. The average area under the ROC curve increased to 0.914.

When we used the sequences of both loops for training and evaluation, the average AUC increased further to 0.933, significantly higher than that of predictors using either V2- ($p < 0.0001$) or V3-sequences ($p = 0.0022$) alone. The differences were even more obvious

when comparing the sensitivities at high specificities (see Figure 3.5). At a specificity of 90%, the predictor using sequence data from both loops improved by 4.5 percentage points in comparison to the predictions based on the V3 loop alone. At the specificity of the 11/25-rule (96.1%) the difference was 5 percentage points.

#### Validation on data derived from clinical isolates

In general, results on the clonal dataset had high quality, not only for predictions derived from sequences of both loops but also for prediction methods trained only on the V3 loop. Similarly accurate V3-prediction methods were also found in previous reports in which prediction methods performed quite well on datasets collected from the Los Alamos Sequence Database. On the other hand, the performance of coreceptor usage prediction from V3-genotype is known to be much worse when tested on clinically derived samples

For example, Low et al. evaluated different publicly available prediction methods on the HOMER cohort of therapy-naïve patients (Low et al., 2007). In this analysis, even the best methods only achieved sensitivities of 50% at a fixed specificity of 90%. An explanation for this decrease in sensitivity might be the fact that the viral population in vivo is a swarm of genetically and phenotypically heterogeneous variants, often termed the quasispecies (Sing et al., 2007), a problem that is not solved by simply using sequence data from another region of the envelope protein. Therefore, it was not clear if the observed improvements in prediction performance on clonal data can be transferred to datasets containing clinically derived samples. We therefore validated our results on two datasets containing such data. The first was a subset of the aforementioned HOMER cohort for which also V2 sequence data was available, while the second was collected at the Virology Institute of Cologne University. This dataset contained therapy-experienced isolates that were screened for administration of maraviroc. With these two datasets we should be able to see possible differences between therapy-naïve and therapy-experienced patients and, furthermore, facilitate direct comparison with previous reports as (a subset of) the same samples as used in the publication by Low et al. could be analyzed here.

**Prediction results on therapy-naïve and therapy-experienced data** We trained the different prediction models used in this analysis with data from the complete clonal dataset. Both evaluations on the two clinical datasets are based on exactly the same prediction models. Due to the inferiority of models using only sequence data from V2 observed in the cross-validation study, we did not use them here. The inputs for the two models, one using sequence information of the V3 loop alone and one using additional information from the V2 loop, comprised the raw sequence information and the position independent features found to be significant (see above). All models were optimized with an internal grid-search on their own. The optimal cost-factors were $c = 0.1$ for the V3-model, and $c = 0.05$ for the combined model (V2/V3).

The Monogram Trofile Tropism Assay (Whitcomb et al., 2007) determined 233 (86.9%) of the 268 antiretroviral naïve samples as being R5 and 35 (13.1%) as CXCR4-capable, respectively. ROC curves displaying the prediction performances in comparison with the Trofile results are depicted in Figure 3.6. The red line indicates the ROC-curve for the method trained solely on V3 data while results of the prediction method trained with

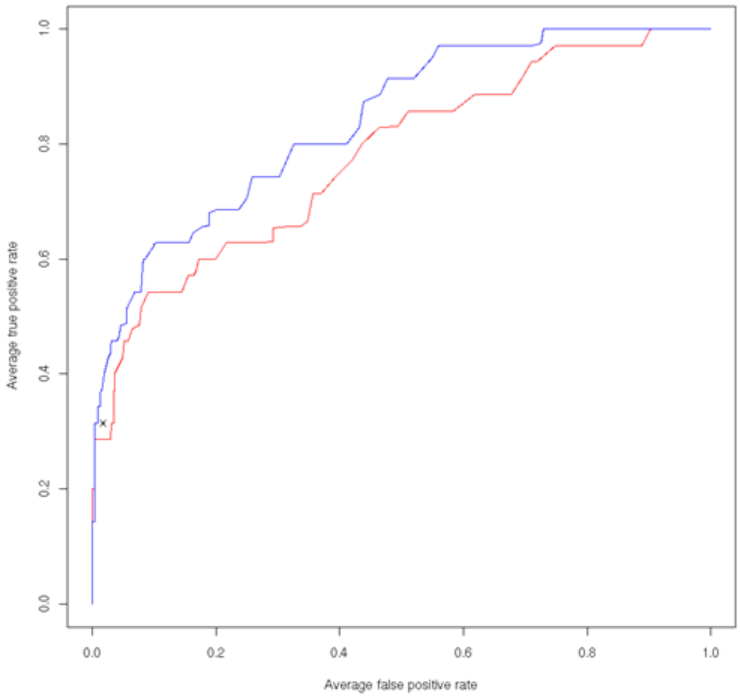


Figure 3.6: Prediction results on the therapy-naïve dataset models. Prediction models were trained on the complete clonal dataset using either sequence information of the V3 loop alone (red line), or using sequence information from the V2 and the V3 loop together (blue line). The x marks the results of the 11/25-rule.

V2- and V3-sequence data are shown in blue. The performance of the 11/25-rule on this dataset is shown by the x in the left part of the figure. It has a very high specificity with 98.3% but the sensitivity only reaches 31.4%. Low et al. found comparable values (30.5% sensitivity and 93.4% specificity) in their analysis of the HOMER cohort (Low et al., 2007). In addition, the results of GENO2PHENO[CORECEPTOR] in this analysis, which was trained on a similar clonal dataset of V3-sequences and also used Support Vector Machines, achieved comparable performances as the V3-model used in our analysis.

As can be seen by the ROC-curves, using sequence information from both loops together yielded better results than models trained on the V3 loop alone. Although these results were still not as good as the results observed on clonal data, sensitivities of the predictor using both loops were higher than for the V3-predictor at all specificities. At a fixed specificity of 90%, there was an increase in sensitivity in 8.6 percentage points (62.8% vs. 54.2%). When comparing the area under the ROC curves, we observed an increase from 0.779 to 0.841.

In the last part of our analysis, we evaluated the same prediction models with the same parameter settings on the dataset of therapy-experienced patients. The fraction of samples capable of using the CXCR4-coreceptor was much higher than on the therapy-naïve dataset. In fact, the Trofile assay determined equal numbers of R5- and X4-samples, reflecting individuals with a higher degree of disease progression. Predictions were better,

in general, than the ones achieved on the HOMER cohort (see Figure 3.7). The 11/25-rule had a sensitivity of detecting X4-viruses of 43.7% while having a specificity of 93.7%. Support Vector Machine models relying entirely on the V3 loop reached a sensitivity of 59.4% at 90% specificity. The incorporation of the V2 loop further increased the sensitivity to 71.8% at the same specificity level. An increase in AUC was also observed from 0.815 of the V3-models to 0.871 for the models using sequence information of both variable loops.

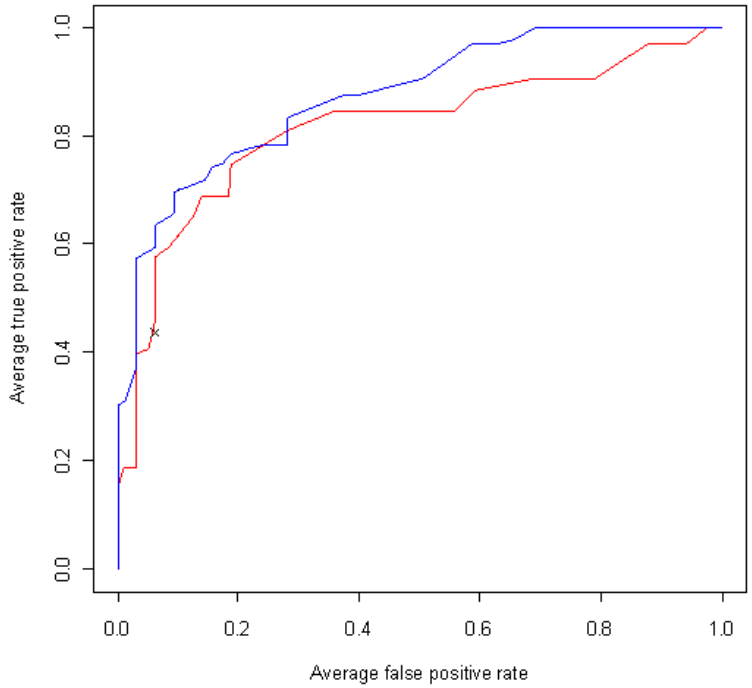


Figure 3.7: Prediction results on the therapy-experienced dataset. Prediction models were trained on the complete clonal dataset and validated on the therapy-naïve dataset were evaluated on a second dataset of 64 therapy-experienced patients. The red line shows the results for the predictor using the V3 loop alone, the blue line describes the ROC-curve of the prediction method using sequence information from the V2 and the V3 loop together. The 11/25-rule is marked by the x.

Last but not least, we compared our results with the two most widely used web-servers for HIV-1 coreceptor usage prediction, both using only V3-sequences as input: geno2pheno[coreceptor] (Lengauer et al., 2007) and WebPSSM (Jensen et al., 2003). geno2pheno[coreceptor] predicted 48.9% of the X4-samples in the HOMER cohort correctly by having a specificity of 90.0%. On the same dataset, WebPSSM displayed 38.3% sensitivity and 94.3% specificity. When analyzing the therapy-experienced samples, both methods performed much better similarly to our results discussed here. Sensitivitiy was 56.2% for both prediction methods, while geno2pheno[coreceptor] achieved 93.7% and WebPSSM 87.5% specificity. Thus, on both datasets the two methods were clearly outperformed by our model using sequence information from both the V2 and the

| | *sensitivity* | *specificity* | *area under the ROC-curve* |
|---|---|---|---|
| **Los Alamos dataset** | | | |
| V2 model | 34.6% | 90.0% | 0.730 |
| V3 model | 85.3% | 90.0% | 0.914 |
| V2 & V3 model | 89.8% | 90.0% | 0.933 |
| 11/25-rule | 66.7% | 96.1% | N/A |
| **Therapy-naïve dataset** | | | |
| V3 model | 54.2% | 90.0% | 0.779 |
| V2 & V3 model | 62.8% | 90.0% | 0.841 |
| 11/25-rule | 31.4% | 98.3% | N/A |
| geno2pheno[coreceptor] | 48.9% | 90.0% | N/A |
| WebPSSM | 38.3% | 94.3% | N/A |
| **Therapy-experienced dataset** | | | |
| V3 model | 59.4% | 90.0% | 0.815 |
| V2 & V3 model | 71.8% | 90.0% | 0.871 |
| 11/25-rule | 43.7% | 93.7% | N/A |
| geno2pheno[coreceptor] | 56.2% | 93.7% | N/A |
| WebPSSM | 56.2% | 87.5% | N/A |

Table 3.3: Overview of performances of different prediction methods on the three datasets analyzed.

V3 loop. All prediction results are summarized in Table 3.3.

### 3.1.3 Discussion

We have shown that several sequence positions in the second hypervariable loop of gp120 are associated with coreceptor usage. In comparison to mutations within V3, the observed mutations in V2 showed lower absolute correlation. However, some mutations (e.g. V169T or the S195H) showed strong correlation with CXCR4-usage, suggesting their use in prediction models.

The question whether these mutations are responsible for a switch in phenotype remains unclear. Some authors proposed that such mutations could compensate for fitness losses caused by mutations in the V3 loop that directly affect viral tropism (see e.g. Pastore et al., 2006).

Incorporation of V2 into prediction models confirmed our assumptions. Models making use of sequence data from both variable loops improved the performance on all tested datasets. On clonal data, on which standard models using V3 loop data only performed already quite well, the incorporation of additional V2 loop mutations significantly outperformed them, both with respect to the overall performance measured in AUC as well as for sensitivities at high specificities. The enhanced models improved there by 4.5 to 5 percentage points.

Furthermore, we found several position-independent features significantly different between R5- and X4-samples. Most differences were found in a region of the loop that allows

for multiple insertions. We found a broad distribution of loop lengths but could not recognize any pattern behind these insertions. It seemed to be quite random, however, the fact that we found features correlated with coreceptor usage indicates that there was a higher-level pattern not necessarily relying on easily identifiable mutations. In previous reports, the ability of increasing the length of the loop has been assumed to be a defense mechanism against detection by the immune system. Together with extensive glycosylation it might serve to protect viruses from neutralizing antibodies (Derdeyn et al., 2004). Several results found that the variable loops are significantly shorter and less glycosylated in early infection of HIV-1 and suggested that this protection shield is not initially required upon transmission into an immunologically-naïve host. Instead, the viruses are outgrown by strains with more compact variable loops (Derdeyn et al., 2004; Chohan et al., 2005; Liu et al., 2008). One could assume that because X4-viruses predominantly appear in later stages of disease, we should see the length of the loop and the number of N-glycosylation sites also to be correlated with X4-viruses. However, we could not confirm this since neither a higher number of N-glycosylation sites, nor the length of the V2 loop were significantly correlated with X4-viruses.

Predictions were generally quite accurate on the clonal dataset, however similar to previous reports by Low et al. (Low et al., 2007) we observed much worse prediction results when we tested on clinically derived isolates. On these, all tested methods showed substantial losses in performance. Nevertheless, we could show that models using both hypervariable loops, V2 and V3, performed best in terms of area under the ROC curve and sensitivity. For a defined specificity of 90%, they increased in sensitivity by about 8 percentage points on therapy-naïve data and about 12 percentage points on therapy-experienced isolates in comparison to V3-models. On both datasets, but especially for the therapy-naïve samples, our results were still not satisfactory, however the impact of mutations in the V2 loop could clearly be demonstrated.

Several reasons might explain the performance losses on clinically derived data. Among them is certainly the inability of standard bulk-sequencers to detect minority variants. Some X4-phenotypes detected by the Trofile assay could arise from minor variants that are below the threshold of detection of standard sequencers. New sequencing technologies such as pyrosequencing that are able to detect minority populations at very low sensitivity levels should therefore further improve prediction outcome. Another advantage of these methods is that they generate clones rather than "bulk" sequences that reflect the consensus of the viral population. Thus, interpretation problems arising from nucleotide mixtures should be minimized with next generation sequencing technologies. This is especially true for the V2 loop and its extensive length variation. If two viral strains have different lengths it becomes very hard to interpret the electropherograms generated by standard sequencing machines.

In the course of this work, a new Trofile assay has become available. While the clinical datasets we analyzed here were phenotyped with the original Trofile assay, the new enhanced Trofile (ESTA) should have a better sensitivity to detect minority populations. On the other side, the specificity of the new assays seems to have become worse so that similar results to the original Trofile assay can be achieved using a lower specificity cutoff (see Section 5.2). However, if this holds true for predictions based on V2 and V3 remains unclear and should be addressed in future.

A shortcoming of this study is that we had only a limited number of genotype-phenotype pairs from the Los Alamos HIV Sequence Database. Although the dataset consisted of more than 900 genotype-phenotype pairs, the number of pairs to be analyzed was much lower due to some patients contributing several sequences. Instead of the 900 pairs, we had only about 350 to train and evaluate our models. Another point is that the large number of possible insertions in the V2 loop makes it very difficult if not impossible to obtain a perfect alignment. We have tried to optimize the alignment by constructing and aligning against a sequence profile that allowed for insertions, however this approach is far from being perfect and different approaches might improve this in future. Last but not least, the quality of genotypes as well as phenotypes varies extensively in the different datasets since a lot of different assays have been used over time. Hence, low-quality genotypes or phenotypes are possible. All these points may impair our analysis but on the other hand they also prove the stability of the results. Using high quality datasets, our results should become even more pronounced in future.

In order to further improve the prediction models we suggest to incorporate clinical features such as $CD4^+$ T cell counts as described in (Sing et al., 2007) and structural properties into the prediction system in future. Structural information of the V3 loop has been shown to improve predictions (Sander et al., 2007), but unfortunately no structural information of the V2 loop is available, so far.

## 3.2 Gp41 - The Wrong Part of the Envelope?

In the previous section we have seen how mutations in the second hypervariable loop of gp120 are correlated with viral tropism and how these mutations can be incorporated to improve HIV-1 coreceptor usage prediction. Of course, the V2 loop is not the only region of the envelope affecting coreceptor phenotype. Mutations in the first hypervariable (V1) loop (Boyd et al., 1993) and the conserved region 4 (C4) (Carrillo and Ratner, 1996) have been shown to be associated with R5- or X4-viruses. Here, however, we want to focus on another region that recently came into focus: the gp41 protein which forms the envelope together with gp120. Huang et al. mapped determinants modulating coreceptor specificity to the gp41 transmembrane (TM) subunit. With site-directed mutagenesis they could show that a single amino-acid substitution in the gp41 is able to confer a dual-tropic phenotype (Huang et al., 2008). Previously, also alterations in the Lentivirus Lytic Peptide Domain 2 (LLP-2) of gp41 have been shown to modulate the syncytium-inducing ability of Env (Kalia et al., 2003).

Intuitively, the gp41 protein seems to be the wrong place to search for determinants affecting HIV-1 coreceptor usage because it binds neither to CD4 nor to the coreceptor itself. Instead, the main task of gp41 is to anchor the glycoprotein complex within the host-derived membrane and to promote fusion of the viral envelope and the target cell membrane (Derdeyn et al., 2000). On the other hand, engagement of the coreceptor is thought to promote additional conformational changes in gp41 (Pancera et al., 2010). Thus, if binding to a different coreceptor influences these changes, the virus has to adapt to it and mutations within gp41 might be a result of the switch in phenotype.

The other direction, that mutations within the V3 loop influence susceptibility to enfuvirtide (T-20), a drug targeting the gp41 protein, has been demonstrated in several works (e.g. Derdeyn et al., 2000, 2001; Reeves et al., 2002). Most mutations were reported in the bridging sheet or the V3 loop reducing affinity for CCR5 and increasing sensitivity to T-20 (Reeves et al., 2002). However, while such an analogy (gp120 mutations affect susceptibility of a drug against gp41, and gp41 mutations change susceptibility of coreceptor antagonists) sounds very attractive, the provided explanations in these publications do not make sense with respect to the direction we are looking at. The authors hypothesized that not structural alterations are the reason for increased susceptibility but suggested kinetic reasons. In order to pursue this line of thought, one has to understand how T-20 works. This drug is a peptide based on the HR2 domain of gp41 and competes with this domain for binding to HR1. Binding of HR1 to HR2 leads to a six-helix bundle whose formation is prevented by T-20. However, the binding domains are only exposed to each other in a short time window that appears to be opened by CD4 binding and closed by coreceptor engagement (Reeves et al., 2002). The more time the virus needs to bind to the coreceptor, the more time has the drug to bind and block the receptor, too. Since a higher affinity to a receptor usually coincides with faster binding it is not surprising that determinants of coreceptor usage have an effect on viral susceptibility to T-20. Thus, direct interaction between gp120 and gp41 does not seem to be necessarily there.

What does this tell us about our issue: can mutations within gp41 alter a virus from R5 to X4 and can we use this information for improving predictions of coreceptor usage? In principle it tells us nothing, but in our opinion it is important to have in mind that

mutations can indirectly modulate the phenotype or, in this case, resistance to a drug, not just by altering the drug target itself but also by mechanisms of which we are not directly thinking and which have no effect on the target region. Therefore, although it does not seem to make sense to search for mutations in regions or proteins that are not in contact with the coreceptors, we should not shy away from analyzing such data.

In this section we have analyzed two large well-defined HIV-1 patient cohorts containing more than 2000 genotype-phenotype pairs for correlations of gp41-sequences with HIV-1 coreceptor usage determined by the Monogram Trofile Tropism Assay (Whitcomb et al., 2007). The aim was to find mutations associated with tropism and to assess the possible additional improvement gained from including gp41 sequence variation into coreceptor usage prediction methods.

The work described in this section was carried out in collaboration with the British Columbia Center for Excellence in HIV (Vancouver, British Columbia, Canada), Pfizer Inc. (New York, USA), and Pfizer Research & Development (Sandwich, UK). It has been presented at the 8th European HIV Drug Resistance Workshop 2010, Sorrento, Italy (Thielen et al., 2010a) and has been accepted for publication in *Antiviral Therapy* under the title "Mutations in gp41 are correlated with coreceptor tropism but do not improve prediction methods substantially".

### 3.2.1 Material & Methods

#### Datasets

We used data from two large datasets that were independent from each other and that contained samples from patients at different stages of disease. One dataset consisted of therapy-naïve isolates while the other dataset was composed of samples from heavily pre-treated patients.

Therapy-naïve samples were derived from the *British Columbia HAART Observational Medical Evaluation and Research cohort (HOMER)*. This cohort includes all HIV-positive, antiretroviral-naïve adults who started HAART though the British Columbia Drug Treatment Program between August 1996 and September 1999 (Hogg et al., 2001). Overall, 1188 patients were enrolled in the cohort but gp41 sequences and a corresponding phenotype were only available for 435 subjects (369 R5 and 66 D/M or X4). Coreceptor usage was determined using the prototype Monogram Trofile Assay.

The second dataset comprised the screening samples from the *Maraviroc versus Optimized Therapy in Viremic Antiretroviral Treatment-Experienced Patients (MOTIVATE)* 1 and 2 studies (studies A4001027 and A4001028, respectively) (Gulick et al., 2008; Fätkenheuer et al., 2008). These were conducted to characterize the efficacy and safety of maraviroc in treatment-experienced patients. Similar to the HOMER cohort, these samples were screened for coreceptor usage with the original Monogram Trofile Assay. Patients screened to have an R5-phenotype were enrolled into the MOTIVATE studies, while a subset screened as having CXCR4-using viruses entered the A4001029 (1029) study. We refer to the entire dataset of those screened for MOTIVATE 1 and 2, including those entering A4001029, as the *MOTIVATE dataset* in this section. Before starting therapy, the gp41 protein of all samples was sequenced for resistance to enfuvirtide (T-20, Fuzeon). Retrospectively, also the V3 loop of gp120 was sequenced for tropism analyses. Altogether,

1915 samples were available with sequence information for both the V3 loop and the gp41 region as well as a Trofile result (1212 R5, 703 D/M or X4 by Trofile, from now on referred to as X4). In addition, for a subset of 859 patients who entered the studies, we have also information on therapy outcome with maraviroc.

### Sequencing

**V3 loop** Viral RNA was extracted from frozen plasma samples using the standard operating procedures for the NucliSENS easyMAG (bioMérieux). The V3 loop of the envelope protein gp120 was amplified from the RNA extracts with nested RT-PCR methods. Employing SuperScript™ III One-Step RT-PCR system with Platinum® Taq High Fidelity enzyme (Invitrogen) the reverse transcriptase (RT) and first-round PCR reactions were combined into one step using forward primer SQV3F1 (5' GAG CCA ATT CCC ATA CAT TAT TGT 3') and reverse primer CO602 (5' GCC CAT AGT GCT TCC TGC TGC TCC CAA GAA CC 3'). The second-round PCR was then performed using the Expand™ High Fidelity enzyme (Roche Diagnostics) with forward primer SQV3F2 (5' TGT GCC CCA GCT GGT TTT GCG AT 3') and reverse primer CD4R (5' TAT AAT TCA CTT CTC CAA TTG TCC 3'). Amplicons taken from the second-round PCR reaction were sequenced in both the 5' and 3' directions using population-based sequencing on an ABI 3730 automated sequencer (Applied Biosystems). The sequencing reaction required the use of BigDye® Terminator (Applied Biosystems) with forward primer V3O2F (5' AAT GTC AGY ACA GTA CAA TGT ACA C 3') and reverse primer SQV3R1 (5' GAA AAA TTC CCT TCC ACA ATT AAA 3').

**gp41** For sequencing of the gp41 protein, viral RNA was extracted from frozen plasma sample using a Guanidine Thiocyanate based lysis buffer followed by precipitation using ispropanol. A single nested RT-PCR amplification was performed for each sample. The RT step was performed from the extracted RNA using Expand RT enzyme (Roche Diagnostics) and the reverse primer GP41RO (5' CTT TTT GAC CAC TTG CCA CCC AT 3'). Subsequently, a first-round PCR amplification was performed using Expand™ High Fidelity enzyme and the forward primer GP41FO (5' TTC AGA CCT GGA GGA GGA GAT AT 3'). Second-round PCR was performed using Expand™ High Fidelity enzyme with forward primer GP41Fi (5' GGA CAA TTG GAG AAG TGA ATT AT 3') and reverse primer GP41Ri (5' CTG TCT TAT TCT TCT AGG TAT GT 3'). Second-round PCR amplicons were subsequently sequenced using an ABI 3700 or 3730xl automated sequencer (Applied Biosystems) using BigDye® Terminator V3.1 (Applied Biosystems) with 11 sequencing primers - 6 forward primers: GP41.1A ( 5' CCA TTA GGA GTA GCA CCC ACC 3'), GP41.2A ( 5' ATC AAG CAG CTC CAG GCA AGA 3'), SQGP41F ( 5' TTT GTG GAA TTG GTT TRA CAT AA 3'), ENV8340F ( 5' AAT AGA GTT AGG CAG GGA TAC TCA CC 3' ), ENV8435F ( 5' TGG AGA GAG AGA CAG AGA CAG ATC C 3'), ENV8522F ( 5' CAG CTA CCA CCG CTT GAG AGA CTT A 3'); and 5 reverse primers: ENVR ( 5' TGC CTG GAG CTG CTT GAT GCC CCA GAC 3'), GP41.1B ( 5' TTG TTC ATT CTT TTC TTG CTG 3'), SQGP41R ( 5' GCC TCC TAC TAT CAT TAT GAA TA 3'), SC90R ( 5' GGA TCT GTC TCT GTC TCT CTC TCC A 3'), SC52R ( 5' TAA GTC TCT CAA GCG GTG GTA 3'). The two second-round PCR primers also

served as back-up sequencing primers.

#### Sequence alignment, encoding and coreceptor usage prediction

Multiple sequence alignments of the V3 loops and the gp41-sequences were generated using ClustalW (Thompson et al., 1994) with standard parameters following manual inspection. The aligned sequences were then encoded for input in prediction methods using a canonical indicator representation. Each position is represented by a 21-bit vector indicating the presence or absence of a specific amino acid. The last bit in this vector represents a gap at the respective position in the aligned sequence. For ambiguous nucleotides, the respective codon was translated into all amino acids possible by this sequence mixture. The bits of the respective amino acids were then all set in the encoded vector.

Coreceptor usage was predicted from genotype using Support Vector Machines (SVMs). Because non-linear kernels did not outcompete linear kernels in previous studies using simple encodings for coreceptor usage prediction (Sing et al., 2007) and because linear kernels are easier to interpret, we employed linear kernels in this work, too. Support Vector Machine models were generated with *LIBSVM* (Chang and Lin, 2001) implemented in the R-package *e1071*. The cost-parameter $c$, controlling for the trade off between allowing training errors and forcing rigid margins, was optimized with an internal grid-search within a range from $c = 10^{-6}$ to 10.

#### Statistical analysis

Correlations of specific mutations with tropism were assessed using Fisher's exact test. Association of pairwise interactions among mutations associated with CXCR4-usage was computed with the Matthews correlation coefficient and statistical significance determined using Fisher's exact test, too. In both cases, significance was corrected for multiple testing with the Benjamini-Hochberg method at a false discovery rate of 5%. Models predicting coreceptor usage were evaluated using Receiver Operating Curves (ROC) analysis in ten-fold cross validation and by testing models from one dataset against the other. As measures, we used the sensitivity to detect X4-viruses, specificity, and the area under the ROC curve (AUC). The statistical programming language R (Team, 2005) with the packages *ROCR*, *covaRius* (Sing et al., 2004, 2005), and *e1071* (Chang and Lin, 2001) was used for all calculations.

### 3.2.2 Results

#### Gp41-mutations associated with HIV-1 coreceptor usage

We assessed every amino acid at every position in the V3 loop and the gp41 protein for correlation with HIV-1 coreceptor usage. The analysis was carried out independently for both the HOMER and the MOTIVATE dataset. Amino acids predictive for either R5- or X4-viruses in the respective datasets are shown in Table 3.4. The table only depicts substitutions that were predictive for one of the two phenotypes in the MOTIVATE dataset, and only these amino acids that remained significant after correction for multiple testing. In the HOMER dataset, we could not find any additional residues significant for R5- or X4-viruses after correction for multiple testing. The prediction model on the HOMER

dataset showed generally less statistical power than that on the MOTIVATE dataset and only a few amino acids in the V3 loop or gp41 remained significant there after correction for multiple testing. This is probably due to the smaller size of the former dataset and because of the fact that the patients in MOTIVATE were heavily treatment-experienced whilst the HOMER cohort is compiled from therapy-naïve patients. Thus, viral evolution has probably not gone as far in the HOMER cohort than in MOTIVATE.

Several amino acids showed strong specificity for X4-viruses in gp41. The most notable difference between R5- and X4-viruses was seen for an insertion after the third amino acid of HXB2-gp41. An insertion of a hydrophobic residue there was highly predictive for CXCR4-using variants in both the MOTIVATE and the HOMER cohort. Not only the displayed insertions of an isoleucine, leucine, methionine, or valine proved to be significant for X4-viruses but also the insertion of an alanine (p=0.0011). However, the alanine failed to meet significance after correction for multiple testing. We also compared R5- and X4-viruses in the two datasets for the presence of any hydrophobic insertion after position 3. In the MOTIVATE dataset, we found 5.7% of the R5-samples and 25.2% of the X4 samples to have such an insertion. Even more pronounced differences were found in the HOMER cohort, in which we observed the insertion in 7.7% of the R5- and 46.3% of X4-samples. For comparison, a positively charged residue at position 11 or 25 of the V3 loop was in this dataset only present in 5.1% of the R5-samples and 36.3% of the X4-samples, i.e. the single insertion of a hydrophobic residue after position 3 in gp41 was more predictive for X4-viruses than the 11/25-rule. However, in MOTIVATE the 11/25-rule performed much better by determining 10.4% of the R5- and 60.1% of the X4-viruses to be X4.

As far as we know, this insertion has never been described before. We therefore checked the Los Alamos HIV Sequence Database for the presence of this specific insertion (February 2010). When considering only at most one R5 and one X4 sequence per patient we found 339 R5- and 267 X4-viruses in total. From these, 2.6% of the R5- and 10.5% of the X4-viruses harbored the insertion (data not shown).

Several other amino acids were also significant for one of the two coreceptors, however, only the mutations A30T and L34M, showed an odds ratio of X4 of more than three in both datasets. Primary resistance mutations for T-20, i.e. mutations at positions 36 to 45, were not correlated with HIV-1 coreceptor usage.

The two datasets showed a high degree of concordance in the N-terminal region of gp41 comprising the N-terminal heptad repeat 1 (HR1). Downstream of HR1, however, major differences were found. For example, while K172K was the second most significant mutation in the MOTIVATE dataset, no significant difference was seen in R5- and X4-samples in the HOMER cohort ($p = 0.54$).

#### Interactions between sites in gp41 and V3 predictive for X4-viruses

In the next part of our analysis, we wanted to check if amino acids in the gp41 protein found to be significantly associated with CXCR4-using viruses are correlated with determinants in the V3 loop predictive for X4-viruses. Such interactions might occur due to viral fitness reasons or as a consequence of steric interference introduced by mutations in the V3 loop which has to be resolved by other mutations within the envelope protein. A covariate analysis was therefore carried out with the most significant amino acids within gp41 and the

| | *MOTIVATE* | | | | *HOMER* | | | |
|---|---|---|---|---|---|---|---|---|
| ***aa*** | ***fR5*** | ***fX4*** | ***OR*** | ***pval*** | ***fR5*** | ***fX4*** | ***OR*** | ***pval*** |
| **A1A** | 92.74% | 97.44% | 1.05 | -5.19 | 94.04% | 96.97% | 1.03 | -0.25 |
| 3T | 16.01% | 7.54% | 0.47 | -7.38 | 8.13% | 3.03% | 0.37 | -0.70 |
| **-3aI** | 0.41% | 4.98% | 12.07 | -10.57 | 0.00% | 7.58% | Inf | -4.15 |
| **-3aL** | 0.08% | 2.42% | 29.31 | -6.38 | 0.00% | 0.00% | N/A | 0.00 |
| **-3aM** | 0.91% | 5.12% | 5.64 | -7.53 | 2.44% | 21.21% | 8.70 | -6.59 |
| **-3aV** | 2.06% | 11.95% | 5.79 | -17.96 | 2.44% | 19.70% | 8.08 | -5.91 |
| -3a- | 78.14% | 63.73% | 0.82 | -10.81 | 73.71% | 46.97% | 0.64 | -4.47 |
| **I4I** | 35.56% | 44.38% | 1.25 | -3.83 | 51.76% | 62.12% | 1.20 | -0.85 |
| **S23A** | 5.03% | 10.10% | 2.01 | -4.31 | 1.90% | 7.58% | 3.99 | -1.63 |
| **A30T** | 1.73% | 10.10% | 5.83 | -15.09 | 0.81% | 12.12% | 14.91 | -4.67 |
| Q32L | 21.37% | 10.95% | 0.51 | -8.43 | 22.76% | 15.15% | 0.67 | -0.71 |
| **Q32Q** | 78.38% | 88.90% | 1.13 | -8.52 | 78.59% | 86.36% | 1.10 | -0.74 |
| L34L | 99.67% | 95.73% | 0.96 | -9.30 | 98.10% | 95.45% | 0.97 | -0.74 |
| **L34M** | 0.91% | 7.11% | 7.84 | -12.65 | 2.17% | 7.58% | 3.49 | -1.48 |
| **N42N** | 78.80% | 85.92% | 1.09 | -4.02 | 91.60% | 96.97% | 1.06 | -0.69 |
| **I69I** | 5.94% | 14.22% | 2.39 | -8.62 | 5.69% | 7.58% | 1.33 | -0.24 |
| A96A | 63.61% | 50.64% | 0.80 | -7.53 | 78.05% | 66.67% | 0.85 | -1.23 |
| **A96T** | 27.15% | 35.99% | 1.33 | -4.24 | 17.07% | 34.85% | 2.04 | -2.67 |
| **N160N** | 74.17% | 84.35% | 1.14 | -6.74 | 85.37% | 89.39% | 1.05 | -0.35 |
| N160S | 30.03% | 21.05% | 0.70 | -4.75 | 16.53% | 15.15% | 0.92 | -0.07 |
| T165S | 44.47% | 33.29% | 0.75 | -5.78 | 22.22% | 21.21% | 0.95 | 0.00 |
| **T165T** | 59.90% | 70.13% | 1.17 | -5.15 | 80.76% | 81.82% | 1.01 | 0.00 |
| **K172K** | 80.36% | 93.74% | 1.17 | -16.22 | 86.99% | 90.91% | 1.05 | -0.27 |
| K172R | 27.97% | 10.81% | 0.39 | -18.98 | 23.58% | 18.18% | 0.77 | -0.37 |
| **V182I** | 77.64% | 85.49% | 1.10 | -4.58 | 57.18% | 56.06% | 0.98 | -0.05 |
| V182V | 31.27% | 20.91% | 0.67 | -6.12 | 52.30% | 51.52% | 0.98 | 0.00 |
| V278V | 3.05% | 0.57% | 0.19 | -3.92 | 0.54% | 1.52% | 2.80 | -0.41 |
| **S293S** | 72.52% | 80.51% | 1.11 | -4.05 | 82.66% | 87.88% | 1.06 | -0.43 |

Table 3.4: Amino acids in gp41 significantly associated with HIV-1 coreceptor usage in the MOTIVATE dataset after correction for multiple testing. Numbering is relative to HXB2. "3a" describes an insertion after the third position of gp41. Residues predictive for X4-viruses are shown in bold. Note that HXB2 is an X4-virus, therefore also the amino acid of the "wild-type" can be predictive for X4-usage (e.g. K172K). aa: amino acid; fR5: frequency in R5-viruses; fX4: frequency in X4-viruses; OR: odds ratio (of X4); pval: p-value (log10)

20 most significant residues within V3. Both sets of amino acids were based on results from the MOTIVATE dataset. If several amino acids were found to be significantly associated with X4-viruses at the same position, we combined them. An exception was position 24 (note that HXB2 has two insertions and one deletion, hence position 25 in HXB2 is 24 in the consensus) which we separated into positively and negatively charged residues.

The results of this analysis are visualized in Figure 3.8. The numbers in the boxes display the Matthews correlation coefficient and the log10 p-value of the significance. Significant interactions are illustrated by framed boxes. It is apparent that only moderate covariance between the residues of the gp41 protein and the V3 loop exist.

The interactions observed between different regions (V3 loop and gp41) were between A30T in gp41 and a negatively charged residue at position 24 ($r = 0.2047$, HXB2-position 25) or a positively charged residue at position 25 ($r = 0.1975$, HXB2-position 26) of the V3 loop. Associations between the insertion of a hydrophobic residue at position three in gp41 and amino acids in the V3 loop were much lower and no clear covariation pattern between residues in V3 and gp41 could be found.

#### HIV-1 coreceptor usage prediction with gp41-mutations

We next tried to assess to which extent gp41-sequence information can help to predict HIV-1 coreceptor usage. Because of the observed differences between the HOMER and the MOTIVATE dataset (see above), we restricted the input for our prediction models to positions at the N-terminal end of gp41 up to the last position of the heptad repeat region 1 for which we found high concordance.

Three different kinds of prediction models were generated in this section: a model trained with sequence information from the V3 loop alone, one with data from gp41 alone, and a third model incorporating sequence information of both regions. First, these prediction models were trained and evaluated on each dataset individually by performing ten-fold cross validation. Figures 3.9 A) and B) show the results for the HOMER and the MOTIVATE datasets, respectively.

On both datasets, predictions based solely on sequence information of the gp41 protein achieved relatively high AUCs, confirming the impact of amino acids found to be associated with phenotype. For the HOMER dataset these models reached an AUC of 0.713 while for MOTIVATE data it was slightly better with 0.736. The value of these predictions based on gp41-data becomes clear when we compare the results on the HOMER dataset with predictions based on the V3 loop. These were not much better with an AUC of 0.773.

When using sequence information of both envelope regions for training and evaluation, the AUC significantly increased (AUC of 0.814, $p < 0.0001$) on the HOMER cohort in comparison with prediction results when using sequence data from one region alone. The comparison of sensitivities to detect CXCR4-using variants at a specificity of 90% did not show major differences between the three models on the HOMER cohort. Considering the prevalence of an insertion of a hydrophobic residue at the N-terminal end of gp41, it was not very surprising that prediction methods trained on gp41-sequences only achieved a sensitivity of 46.9%. Sensitivity of V3-models reached 54.5% which is in line with previous reports analyzing a larger subset of the original HOMER cohort (see Low et al., 2007, and Section 3.1). However, we would have expected better results from predictions using both

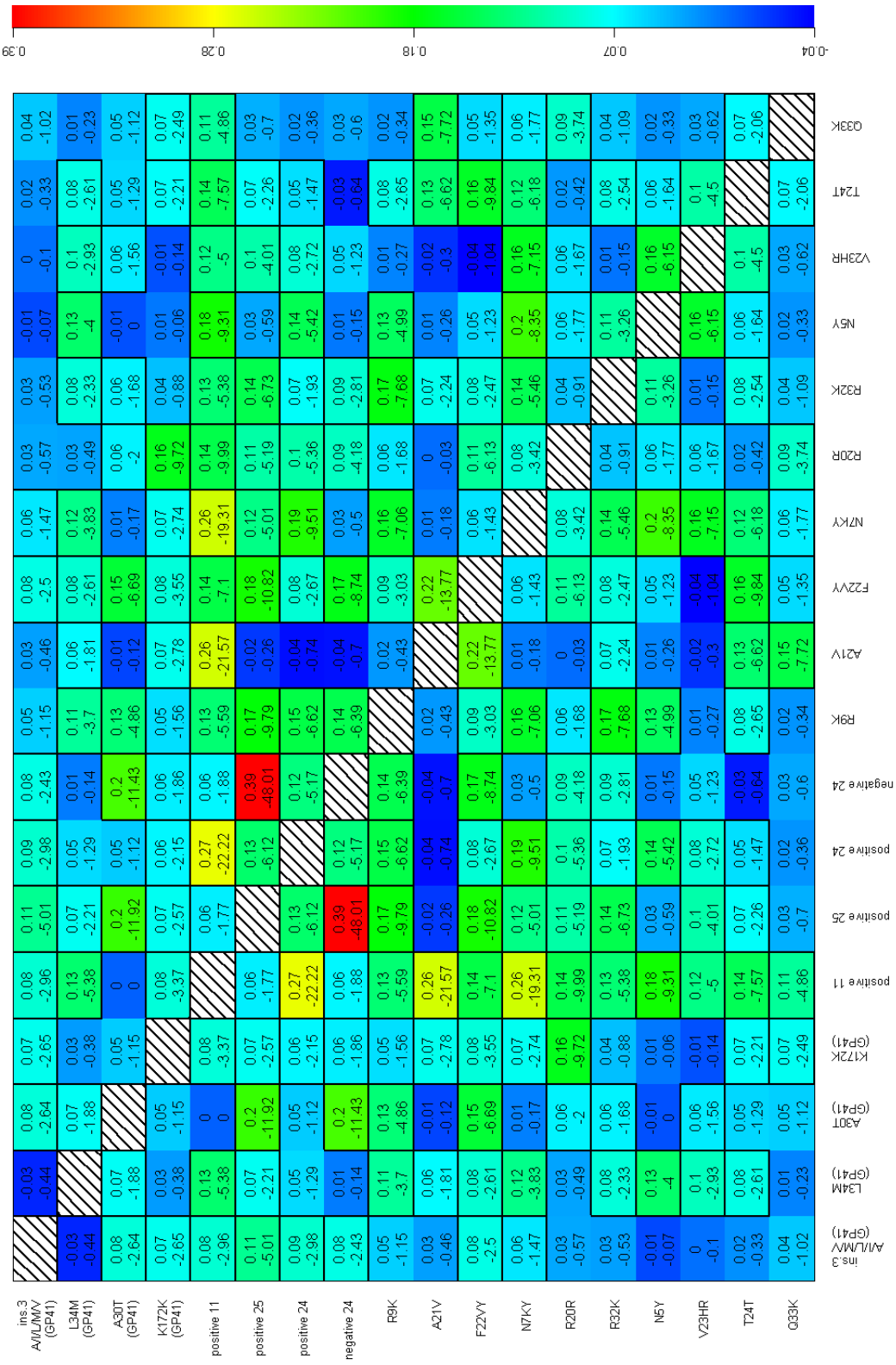


Figure 3.8: Covariance between amino acids predictive for X4-viruses in V3 and gp41. Each box indicates the Matthews correlation coefficient (first number) and the logarithmic (log10) significance (second number) between two positions given by the row and the column.

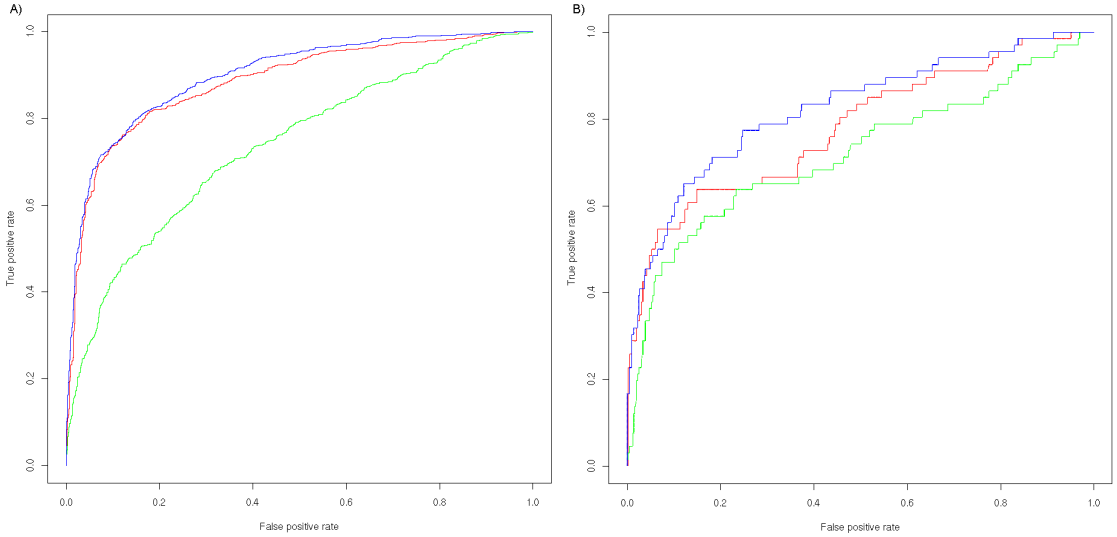


Figure 3.9: ROC-curves of prediction methods trained on sequence information of gp41 (green line), V3 (red line), or both regions together (blue line). A) results from cross validation on MOTIVATE data, B) results from cross validation on HOMER data.

regions as input. In this case, the sensitivity only increased marginally to 57.5%. When analyzing the sensitivities at another specificity-level of 80%, we found more pronounced differences between the different prediction models. There, sensitivity of gp41-models reached 57.5%, the corresponding V3-models 63.6%, while the combined predictor achieved a sensitivity of 71.2% for predicting X4-viruses.

As already mentioned, gp41-models also showed very good results on data derived from the MOTIVATE screening population. However, V3-models were much better than on the HOMER cohort and clearly outperformed the corresponding gp41-models by achieving an area under the ROC curve of 0.884. There was also no substantial improvement over these models when we used sequence information of both regions. AUC increased only marginally to 0.902 which was nevertheless significantly better than predictions based on the V3 loop alone ($p = 0.0002$). The comparison of sensitivities at different specificities yielded similar results. For a defined specificity of 90%, the sensitivity to detect X4-viruses was 42.8% for predictions based solely on gp41, 73.3% for V3-models, and 73.8% for predictions incorporating both the V3 loop and gp41. The corresponding sensitivities at 80% specificity were 54.3% (gp41), 82% (V3), and 82.7% (V3 & gp41).

We next wanted to assess the generalization of our prediction results by training the prediction models on one dataset and validating on samples from the other (and vice versa). Figure 3.10 shows the results when the MOTIVATE dataset was used for training and the HOMER cohort for validation. Concordant with the results obtained by cross validation, predictions based on either gp41-sequence information or the V3 loop alone were comparable to each other with AUCs of 0.774 and 0.811, respectively. However, prediction models trained on data from both regions had significantly higher AUCs (0.881, $p < 0.0001$). The sensitivity at 90% specificity also increased greatly when the two regions were combined, from 51.5% of gp41-models, and 62.1% of models based on the V3 loop,

to 71.2% when predictions were derived from sequence data of both regions. On the other side, just like already seen in the cross-validation analysis, gp41 had only little impact when models were trained on the HOMER cohort and evaluated on the samples of the MOTIVATE screening population. There, predictions based on gp41 alone achieved an AUC of 0.690 which was significantly lower than predictions based on the V3 loop (AUC of 0.842). When gp41 was incorporated into these models, there was only a slight increase in AUC to 0.859. There was also no significant effect seen in terms of sensitivity at 90% specificity. Gp41-models reached a relatively high sensitivity of 37%, but V3-models clearly outperformed them by having a sensitivity of 63.3% and only a minimal improvement to 63.6% was seen when both regions were used as input for prediction methods.

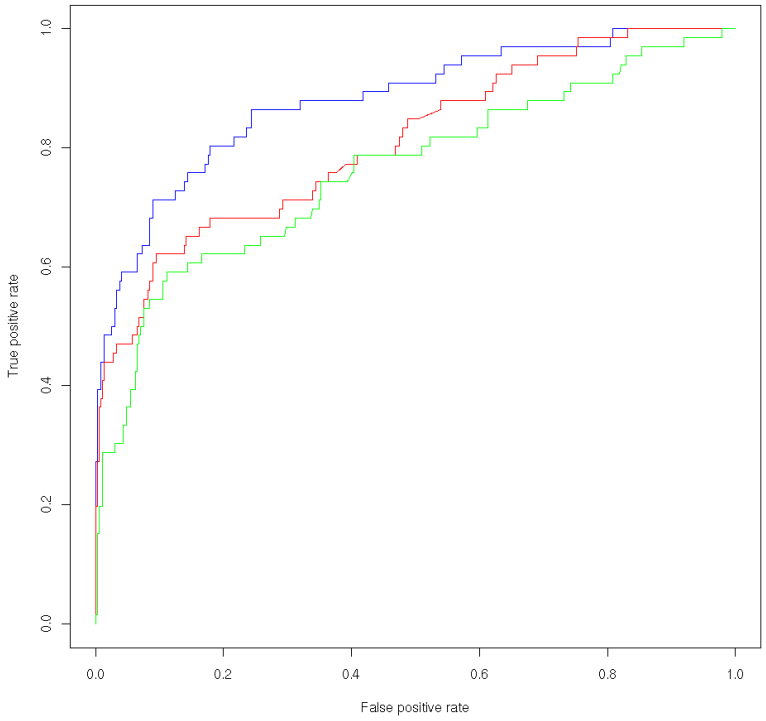


Figure 3.10: ROC-curves of prediction methods trained on sequence information of gp41 (green line), V3 (red line), or both regions together (blue line). Prediction methods trained on MOTIVATE data and validated on HOMER results.

### Prediction of therapy outcome

In the last part of this section, we wanted to assess whether gp41 can help in predicting therapy-outcome of maraviroc based regimens. Therefore, we trained models for HIV-1 coreceptor usage prediction with all samples from the two datasets as described previously. Based on previous works, we selected a cutoff of 95% specificity to predict viruses as R5 or X4.

When comparing the numbers of predicted X4-viruses, we only found small differences between the three prediction models and the results by the Monogram Trofile Tropism

Assay. The phenotypic assay determined 105 of the 852 samples with available treatment outcome to be CXCR4-using. The corresponding numbers for the prediction models were 68 (gp41), 96 (V3), and 99 (V3 & gp41).

Clinical outcome was measured in terms of log10 viral load decrease (*virologic outcome*) and increase in $CD4^+$ cell counts (*immunologic outcome*). Overall, only small differences were seen for the different methods when virological outcome was analyzed. Figure 3.11 depicts the median log viral load decreases of patients carrying R5- (solid lines) or X4-viruses (dotted lines) determined by the different methods within the first 24 weeks of treatment with maraviroc. It is apparent that no major differences between the different methods exist and that predictions based on one region alone seem to perform even slightly better than the Trofile assay and the predictor generated with sequence information from both the V3 loop and gp41. The difference in median log viral load between the two patient groups at week 24 was 1.59 for assignments by the gp41-models (decrease in median log viral load of 2.42 for patients with R5-viruses and 0.83 for patients with X4-viruses), 1.65 (2.45 vs. 0.79) for models trained on the V3 loop alone, and 1.58 for coreceptor usage determined by predictions incorporating both envelope regions (2.44 vs. 0.86). The corresponding values for samples determined to be R5 or D/M by Trofile were quite similar with 2.46 and 0.81.

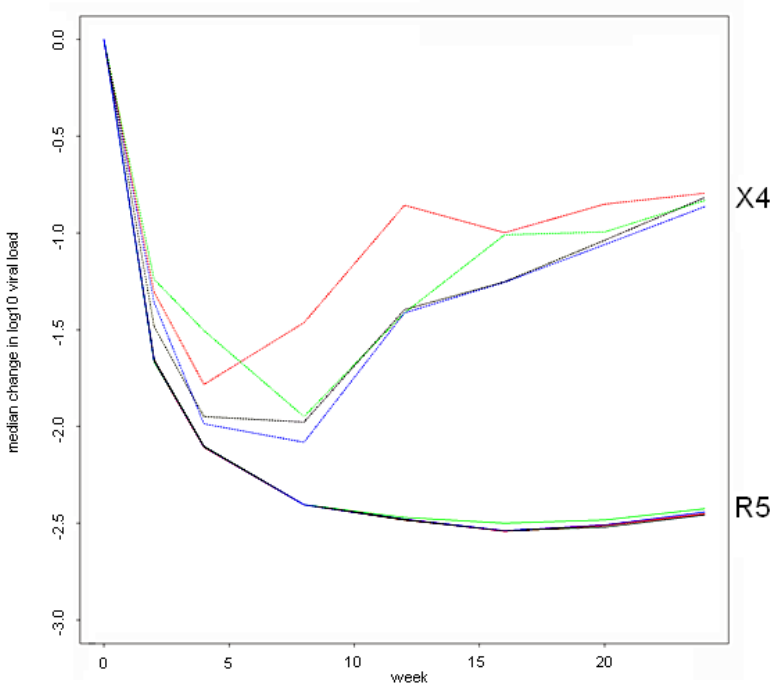


Figure 3.11: Virologic outcomes to week 24. Median log viral load changes in patients determined to be R5 or X4 by different methods. Predictions were done with methods trained for coreceptor usage prediction with a false-positive rate of 5%. Upper lines: X4-predictions, lower lines: R5-predictions, black: Trofile, red: V3-prediction, green: gp41-predictions, blue: predictions derived from models trained on both regions.

The analysis of immunologic outcome yielded comparable results. The increase in $CD4^+$

cell counts was very similar for patients determined to have R5-viruses by the different methods (Trofile +88.5, gp41-predictions +87, V3-models +88.5, predictions using both V3 and gp41 +88.5), however, the immunological outcome of patients predicted to have CXCR4-using viruses was worse when predictions originated by the models trained on the V3 loop only. There, only an increase of 33 $CD4^+$ cells was seen, whereas patients predicted to have an X4-virus by gp41-models showed a median increase of 51 cells, X4-predictions by the combined model 45 cells, and Trofile-D/M patients 47.5 cells.

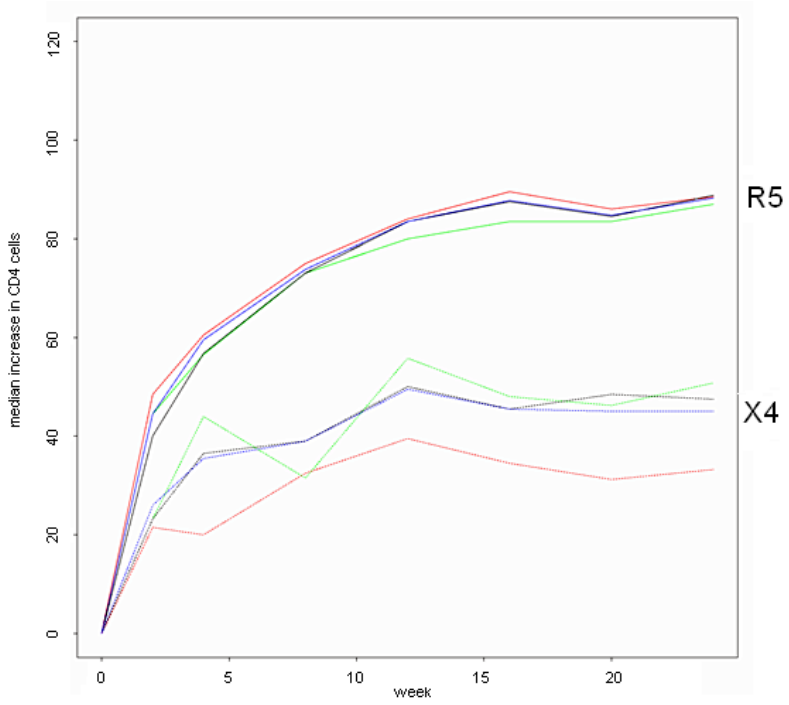


Figure 3.12: Immunologic outcomes to week 24. Median increases in $CD4^+$ cell counts in patients determined to be R5 or X4 by different methods. Predictions were done with methods trained for coreceptor usage prediction with a false-positive rate of 5%. Upper lines: R5-predictions, lower lines: X4-predictions, black: Trofile, red: V3-prediction, green: gp41-predictions, blue: predictions derived from models trained on both regions.

### 3.2.3 Discussion

In this section, we have analyzed the influence of the gp41-glycoprotein for HIV-1 coreceptor usage. We found several mutations to be significantly correlated with CCR5- or CXCR4-using viruses in two independent well-defined cohorts containing more than 2000 samples. An insertion of a hydrophobic residue at the N-terminal end of gp41 was strongly predictive for X4-viruses and even more accurate than the 11/25-rule on the therapy-naïve dataset of the British Columbia HOMER cohort. To our surprise, we could not find any publication describing this insertion. However, at a recent workshop (the XVIII International HIV Drug Resistance Workshop in Fort Myers, Florida, June 2009), Monogram Biosciences presented a list of changes in gp41 to have an impact on coreceptor tropism.

This list of mutations was based on a dataset of 6453 commercially screened subtype B samples that had been run with the original Trofile assay and which had gp41 sequence information. While the aforementioned insertion was surprisingly lacking on this list, the two other mutations we found to be highly significant for X4-viruses, A30T and L34M, were associated with higher infectivity in CXCR4 cells and lower infectivity in CCR5 cells in this work, too.

The basis of the effects of these mutations still remains unclear. It might be that they just compensate for structural rearrangements due to mutations in the gp120 protein. On the other side, Blish et al. have identified two mutations in gp41 that expose conserved Env regions which are desired targets for protective antibodies to HIV-1 (Blish et al., 2008). Of course, the ability to induce these structural changes might not only affect regions targeted by the immune system but could also alter the coreceptor binding site and the preference for a different coreceptor. Interesting in this context is that one of these mutations I675V is adjacent to amino acids we found here predictive for R5- and X4-viruses: T1165S and T165T (HXB2-position 676). Another report also pointed to this position where I675M and N668S were found to enhance sensitivity to soluble CD4 (sCD4), and more interestingly anti-V3 antibodies (Back et al., 1993). Thus, although these residues were not as significantly correlated as others presented here, a link between them and coreceptor usage determinants seems to be possible, too.

Although the discovery of residues in gp41 specific for the two phenotypes is potentially useful for predicting HIV-1 coreceptor usage, we could not demonstrate this consistently in our analyses. For the HOMER cohort, major improvements were obvious. The sensitivity to detect CXCR4-using variants reached more than 70% for a specificity of 90%. In comparison to predictions based on the V3 loop alone and previous works (Low et al., 2007) on this dataset which both reported only a sensitivity of about 50%, this is a dramatic increase in prediction performance. However, we could not observe such an impact on data derived from the MOTIVATE screening population, neither when analyzing sensitivity to detect X4-viruses, nor when the area under the ROC curve or virological outcome was used as measure. In contrast, for reasons that remain unclear, prediction models using only the V3 loop, seemed to be better when analyzing clinical outcome, especially when predicting the increase in $CD4^+$ cell counts.

A reason for the missing effect on the MOTIVATE dataset might be the smaller training set of HOMER samples. It could be argued that using more samples for training, the performance of predictions based on both regions would also increase significantly. However, in our opinion this would not happen because in the cross-validation analysis on the MOTIVATE dataset, the training set in each fold was based on about 1700 samples, either. And in this analysis we also did not encounter substantial improvements of prediction quality. Another explanation might be that the HOMER cohort consists of therapy-naïve patients whereas the MOTIVATE screening population contained heavily pretreated patients. Although it is possible that previous exposure to antiretroviral drugs has an influence on HIV-1 coreceptor usage and that the appearance of specific residues is associated with one of the two phenotypes, we would have expected that these mutations should be more helpful in late stage of disease and not on therapy-naïve data.

Limitations of this analysis certainly include the use of conventional "bulk"-sequencers to determine genotypes and the original Trofile assay for phenotyping. In both cases,

minority populations in the viral quasispecies might have been overseen. A solution could be the use of deep-sequencing technologies and the new improved enhanced sensitivity Trofile assays (ESTA) which are both reported to be able to detect variants at very low minorities.

In summary we conclude that because we could not demonstrate substantial improvements of prediction methods by integrating gp41-sequence information on the MOTIVATE dataset, we believe that, at present, sequencing this part of the envelope protein is not necessary in daily routine when patients are screened for administration of coreceptor antagonists. The additional effort of sequencing the approximately 1035 base pairs of gp41 compared to 105 bp for the V3 loop is not justified by the marginal improvements observed. However, the high correlation of the insertion we found to be predictive for X4-viruses as well as other amino acids highly correlated with phenotype suggests further investigation of gp41 in the realm of HIV-1 coreceptor usage.

## 3.3 Other Parts in the Envelope and Other Genes

In the previous sections we have analyzed two parts of the viral envelope protein for associations with HIV-1 coreceptor usage. Here, we want to discuss the remaining regions as well. Only a handful of mutations in regions other than V2, V3, and gp41, have been found to be correlated with tropism in previous works. The probably best known is a mutation in the conserved region 4 of the gp120 glycoprotein (Carrillo and Ratner, 1996; Rosen et al., 2008).

However, in contrast to the V2 and the V3 loop, as well as the gp41 protein, we do not have sufficient clinical data for validation. Thus, we will mainly restrict ourselves to a descriptive analysis. Our goal will be to find positions that could play a role for tropism which should later on be validated on clinical datasets or with in vitro experiments.

Viral cell entry is mediated mainly by the envelope proteins gp120 and gp41, the extracellular determinants of the virus interacting with the CD4-receptor and the chemokine coreceptors. Thus, one might think that it does not make sense to analyze intracellular proteins of HIV with respect to coreceptor usage. On the other hand, as already mentioned in the previous section, sometimes we see effects we cannot understand at first sight and which have an indirect influence on coreceptor usage.

With respect to intracellular proteins, such explanations might be more complicated to find but some hints of a potential impact already exist. For example, it is known that the viral enzyme *vpu* promotes virus release by counteracting *tetherin*, a restriction factor that can be induced by interferon-alpha (Neil et al., 2007). This factor "tethers" nascent virions to cell membranes so that they cannot be released efficiently.

Tetherin is highly expressed on the surface of macrophages and dendritic cells which both become HIV-1 infected via the CCR5 coreceptor. In a recent study Schindler et al. could show that a mutation in vpu impaired the replicative capacity of HIV-1 in macrophages as severely as the complete lack of vpu function. Thus, the authors speculated that the ability of vpu to counteract tetherin is an important determinant for HIV-1 cell tropism (Schindler et al., 2010). As a consequence of this scenario viruses with such mutations might be found predominantly in CXCR4-using isolates and less in R5-viruses although there is not a direct link to cell entry.

One could think about other reasons for an effect of intracellular enzymes on tropism. However, irrespective of an explanation of how mutations within these enzymes might affect coreceptor usage, amino acids in specific genes of the viral genome like the polymerase enzymes PR, RT, and INT might be very valuable because these are sequenced regularly and could help in predicting tropism without implying any additional testing costs. So, already for practical reasons it makes sense to analyze this kind of data.

### 3.3.1 Material & Methods

#### Datasets

Genotype-phenotype pairs were downloaded from the Los Alamos HIV Sequence Database (April 2010). Sequence data was derived from the following regions: the variable loops 1, 4, and 5 (V1, V4, V5) and the conserved regions 1-5 (C1-C5) of the envelope protein gp120, the gene encoding for the structural proteins group-specific antigen (gag), and the viral

enzymes protease (PR), reverse transcriptase (RT), integrase (INT), and last but not least the regulatory proteins negative regulatory factor (nef), regulator of virion (rev, two splice variants rev-1 and rev-2), Rnase H (RnaseH), rev response element (rre), trans-activating responsive (TAR) element, trans-activator of transcription (tat, two splice variants tat-1 and tat-2), viral infectivity factor (vif), viral protein R (vpr), and viral protein U (vpu) (see Table 3.5).

| | ***Los Alamos*** | | | ***HOMER*** | | |
|---|---|---|---|---|---|---|
| ***region*** | ***pairs*** | ***#R5*** | ***#X4*** | ***pairs*** | ***#R5*** | ***#X4*** |
| C1 | 1614 | 368 | 105 | 0 | 0 | 0 |
| C2 | 2075 | 426 | 132 | 0 | 0 | 0 |
| C3 | 2859 | 567 | 163 | 0 | 0 | 0 |
| C4 | 2207 | 500 | 160 | 0 | 0 | 0 |
| C5 | 1811 | 461 | 124 | 0 | 0 | 0 |
| C1a | 0 | 0 | 0 | 263 | 215 | 38 |
| V1a | 0 | 0 | 0 | 115 | 96 | 19 |
| C2a | 0 | 0 | 0 | 315 | 269 | 46 |
| C2b | 0 | 0 | 0 | 314 | 269 | 45 |
| C2c | 0 | 0 | 0 | 512 | 428 | 84 |
| C3a | 0 | 0 | 0 | 456 | 389 | 67 |
| C3b | 0 | 0 | 0 | 336 | 281 | 55 |
| V1 | 2136 | 432 | 137 | 0 | 0 | 0 |
| V4 | 2255 | 542 | 161 | 0 | 0 | 0 |
| V5 | 2164 | 496 | 155 | 0 | 0 | 0 |
| PR | 406 | 143 | 37 | 701 | 584 | 117 |
| RT | 180 | 128 | 35 | 525 | 443 | 82 |
| INT | 163 | 113 | 35 | 710 | 598 | 112 |
| gag | 131 | 78 | 26 | 695 | 586 | 109 |
| nef | 406 | 121 | 42 | 916 | 759 | 157 |
| rev-1 | 176 | 119 | 35 | 0 | 0 | 0 |
| rev-2 | 1669 | 388 | 114 | 0 | 0 | 0 |
| RnaseH | 180 | 128 | 35 | 680 | 570 | 110 |
| rre | 1734 | 400 | 116 | 0 | 0 | 0 |
| TAR | 67 | 42 | 19 | 0 | 0 | 0 |
| tat-1 | 173 | 117 | 35 | 518 | 421 | 97 |
| tat-2 | 1736 | 400 | 116 | 0 | 0 | 0 |
| vif | 161 | 112 | 35 | 558 | 451 | 107 |
| vpr | 161 | 112 | 35 | 556 | 451 | 105 |
| vpu | 211 | 151 | 35 | 518 | 421 | 97 |

Table 3.5: Genotype-phenotype pairs of different regions available from the Los Alamos HIV Sequence Database and the HOMER cohort. Rev-1 and -2 and Tat-1 and 2 denote different splice variants. The numbers of R5- and X4-sequences are based on using one sequence per patient.

For the genes PR, RT, INT, gag, nef, RnaseH, tat-1, vif, vpr, and vpu also clinically derived samples with coreceptor usage information were available. These isolates were derived from the British Columbia HAART Observational Medical Evaluation and Research (HOMER) cohort (Hogg et al., 2001). Genotypes were determined as described in Section 3.1 and coreceptor usage determined with the original Monogram Trofile Tropism Assay (Whitcomb et al., 2007).

In addition, sequence information from different parts of the envelope protein of HOMER-patients was available. In general, the region from position 85 of HXB2-C1 to position 51 of HXB2-C3 was sequenced. However, this was done using a set of amplicons starting at different points. Due to difficulties inherent in performing population sequencing on the envelope the final sequences are fragmented and contain chunks of data interspersed with missing parts. Thus, for different positions we have different numbers of samples available. Therefore, instead of the conventional classification into the variable and conserved regions, we use here a different notation: *C1a* refers to positions 85 to 120 in HXB2-C1 while *V1a* encompasses the last 10 positions of HXB2-C1 and the V1-loop. *C2a* are the last 4 positions of the V2 loop and the first 32 of the conserved region 2, *C2b* the next 36 positions of this region, and *C2c* the last 31 positions of C2. Finally, *C3a* covers just the first 5 positions of C3 and *C3b* denotes positions 6 to 40 of HXb2-C3. The numbers of available genotype-phenotype pairs are also shown in Table 3.5.

#### Sequence Analysis

All regions were aligned using ClustalW (Thompson et al., 1994) with standard parameters. Sequences of the envelope regions were aligned with sequence to profile alignments while the sequences of the other regions were aligned to each other with normal multiple sequence alignments.

In case of positions with nucleotide ambiguities in the clinical isolates, the respective codons were translated into all combinations of possible amino acids. Each amino acid was then analyzed independently. For example, in a codon *ARM* (IUPAC notation), the letter "R" and "M" stand for the possible nucleotides A or G, and A or C respectively. Thus, the codons AAA, AAC, AGA, and AGC are possible which encode for the amino acids lysine (K), asparagine (N), arginine (R), and serine (S).

In addition to the association of specific residues with phenotype, we also tested for differences in other properties: the number of potential N-glycosylation sites, the length of the sequences, the number of hydrophobic, polar, small, tiny, aliphatic, aromatic, positively charged, negatively charged, and charged residues, and the number of prolines (all according to Livingstone and Barton, 1993, except for gaps which were removed before).

#### Statistical analysis

The association between specific amino acids and HIV-1 coreceptor usage was assessed using Fisher's exact test. Only residues appearing in at least 5% of either R5- or X4-sequences were analyzed.

Differences in medians of numerical quantities of specific properties were assessed using Wilcoxon's rank sum test. In both cases, correction for multiple testing was performed using the Holm-Bonferroni method. We did not use the Benjamini-Hochberg method in

this context because it is highly dependent on the number of tests performed. Some of the sequences analyzed here are very long (e.g. *gag*-sequences have a length of about 500 residues) implying that the p-value for a mutation in these would have to be much lower to be accepted by multiple testing than the p-value for a corresponding mutation in another smaller region (e.g. the conserved region 4 has a length of 41 amino acids).

### 3.3.2 Results

#### Amino acids associated with viral tropism

In total we found 669 amino acids significantly associated with one of the two phenotypes in the samples from the Los Alamos HIV Sequence Database. 468 of these were correlated with phenotypes in non-envelope proteins whereas 201 mutations belonged to gp120. However, after correction for multiple testing with the Holm-Bonferroni method, only 14 amino acids remained significant. Of these, eight were located within gp120, two in the *gag*-protein, two in the *reverse transcriptase*, and one in *vif* and the *rev response element*, respectively.

A similar pattern was seen for samples derived from the HOMER cohort. 795 amino acids were found to be correlated with either of the two phenotypes in non-envelope proteins. Only 48 were significantly associated with R5- or X4-strains in the subregions of gp120. However, while only 2 *gag*-mutations remained significant after correction for multiple testing, this was found for 5 of the 48 mutations located in the envelope.

We could not identify a clear trend towards mutations in a certain region of gp120 or a certain protein. Instead, the length of the different regions was fairly highly correlated with the number of residues significantly associated with phenotype ($r = 0.656$ for mutations significant for R5- or X4-viruses before correction for multiple testing and $r = 0.542$ for positions significant after Holm-Bonferroni).

Among the amino acids being significantly associated with R5- or X4-viruses after correction for multiple testing (see Table 3.7) were some which have been described before. E.g. it is known that a negatively charged residue at gp120-position 440 (position 22 in Table 3.7) is correlated with positively charged residues at position 25 of the V3 loop and thus also with X4-viruses (Rosen et al., 2008). Consequently, it was not especially surprising to see that a positively charged residue at position 440 (22R) was associated with R5-viruses. Again, this result has also been described before (Hoffman et al., 2002), however aside from these two mutations we could not find any other in the literature. The four mutations found in the gag gene showed very remarkable predictivness for X4-viruses. 58R and 62G were found in more than 50% of the X4-viruses in the Los Alamos dataset but only in about 18% and 8% of the respective R5-sequences. Especially, the mutation 62G showed a predictive power which is usually only achieved by the 11/25 mutations of the V3 loop. Considering that the 11/25 rule does not work very well for samples of the HOMER cohort (see also Sections 3.1 and 3.2) the two gag-mutations in the HOMER dataset also demonstrated comparable performances. A "46/75I"-rule would achieve a sensitivity of 28.4% at a specificity of 90.5% on HOMER-samples. We therefore checked if these 31 X4-sequences had known X4-mutations in the V3 loop but could identify these in only 12 isolates (data not shown).

Within the reverse transcriptase, we found in the Los Alamos dataset lysine and arginine

| | *Los Alamos* | | *HOMER* | |
|---|---|---|---|---|
| ***region*** | ***# sign.*** | ***# after H.-B.*** | ***# sign.*** | ***# after H.-B.*** |
| C1 | 52 | 3 | N/A | N/A |
| C2 | 23 | 0 | N/A | N/A |
| C3 | 42 | 2 | N/A | N/A |
| C4 | 17 | 2 | N/A | N/A |
| C5 | 17 | 1 | N/A | N/A |
| V1 | 25 | 0 | N/A | N/A |
| V4 | 21 | 0 | N/A | N/A |
| V5 | 4 | 0 | N/A | N/A |
| C1a | N/A | N/A | 11 | 0 |
| V1a | N/A | N/A | 7 | 0 |
| C2a | N/A | N/A | 8 | 0 |
| C2b | N/A | N/A | 0 | 0 |
| C2c | N/A | N/A | 7 | 2 |
| C3a | N/A | N/A | 0 | 0 |
| C3b | N/A | N/A | 15 | 3 |
| PR | 11 | 0 | 8 | 0 |
| RT | 34 | 2 | 15 | 0 |
| INT | 17 | 0 | 22 | 0 |
| gag | 140 | 2 | 350 | 2 |
| nef | 32 | 0 | 310 | 0 |
| rev-1 | 6 | 0 | N/A | N/A |
| rev-2 | 23 | 0 | N/A | N/A |
| RnaseH | 11 | 0 | 7 | 0 |
| rre | 55 | 1 | N/A | N/A |
| TAR | 14 | 0 | N/A | N/A |
| tat-1 | 9 | 0 | 16 | 0 |
| tat-2 | 11 | 0 | N/A | N/A |
| vif | 39 | 1 | 24 | 0 |
| vpr | 7 | 0 | 14 | 0 |
| vpu | 59 | 0 | 29 | 0 |
| total | 669 (201) | 14 (8) | 843 (48) | 7 (5) |

Table 3.6: Numbers of amino acids found to be significantly associated with R5- or X4-viruses. The numbers in the brackets in the last row depict the numbers of significant residues in the gp120 regions. Abbreviations: sign.: significant with $p < 0.05$; H.-B.: significant after correction for multiple testing with Holm-Bonferroni.

at position 356 to be correlated with R5- and X4-viruses, respectively. Because we did not get a similar result in the HOMER cohort (see also Figure 3.13) collected from antiretroviral naïve patients, we first assumed an impact of HAART therapy. However, we could not find any NRTI or NNRTI resistance associated with these amino acids.

| *region* | *aa* | *fR5* | *fX4* | *p-value* |
|---|---|---|---|---|
| | ***Los Alamos HIV Sequence Database*** | | | |
| C1 | 23b-noIns. | 0.27% | 7.61% | $3.49 \cdot 10^{-5}$ |
| C1 | 28del | 0.81% | 9.52% | $3.16 \cdot 10^{-5}$ |
| C1 | 29c-noIns. | 99.18% | 89.52% | $8.26 \cdot 10^{-6}$ |
| C3 | 27noIns. | 89.77% | 78.52% | 0.0002 |
| C3 | 29K | 13.58% | 27.60% | $6.83 \cdot 10^{-5}$ |
| C4 | 22E | 10.40% | 26.87% | $1.05 \cdot 10^{-6}$ |
| C4 | 22R | 26.80% | 8.75% | $6.57 \cdot 10^{-7}$ |
| C5 | 25L | 7.15% | 18.54% | 0.0004 |
| gag | 58R | 17.94% | 65.38% | $1.27 \cdot 10^{-5}$ |
| gag | 62G | 7.69% | 57.69% | $4.11 \cdot 10^{-7}$ |
| RT | 356K | 71.87% | 28.57% | $1.27 \cdot 10^{-5}$ |
| RT | 356R | 27.34% | 71.42% | $4.11 \cdot 10^{-7}$ |
| vif | 22K | 19.64% | 54.28% | 0.0001 |
| rre | 95V | 0.00% | 5.17% | 0.0001 |
| | ***HOMER*** | | | |
| C2c | 75M | 0.23% | 7.14% | 0.0001 |
| C2c | 75V | 91.82% | 77.38% | 0.0003 |
| C3b | 26K | 16.37% | 41.81% | $7.69 \cdot 10^{-5}$ |
| C3b | 26N | 79.71% | 49.09% | $6.54 \cdot 10^{-6}$ |
| C3b | 27Q | 60.49% | 30.90% | $8.78 \cdot 10^{-5}$ |
| gag | 46I | 6.48% | 21.10% | $8.60 \cdot 10^{-6}$ |
| gag | 75I | 4.26% | 15.59% | $5.45 \cdot 10^{-5}$ |

Table 3.7: Genotype-phenotype pairs of different regions significantly associated with coreceptor usage after correction for multiple testing. Numbering according to HXB2. Position numbers of C2c and C3b are with respect to the usual regions C2, and C3. Abbreviations: *aa* = amino acid with position, *fR5 / fX4* = frequency of occurring in R5- and X4-viruses, respectively, *a/b/c* = mutual insertions after the displayed HXB2-position, *noIns.* = no insertion, *del* = deletion

**Comparison between clonal and clinically derived samples** What sticks out is that both on clonal as well as on clinically derived data many mutations with an uncorrected p-value of less than 5% were found in the gag gene. Furthermore, after adjusting for multiple testing, in both cases two amino acids remained signficant. However, upon closer inspection of these residues we had to see that, first, they differed, and second, substitutions significant in one dataset did not show a substantial trend toward the same coreceptor in the other dataset (see Table 3.7).

Figure 3.13 shows the frequencies of the significant amino acids in the two datasets. Green points mark the frequencies within R5-samples whereas the frequencies in X4-viruses are shown in red. The x-axis depicts the frequency of a substitution in the Los Alamos

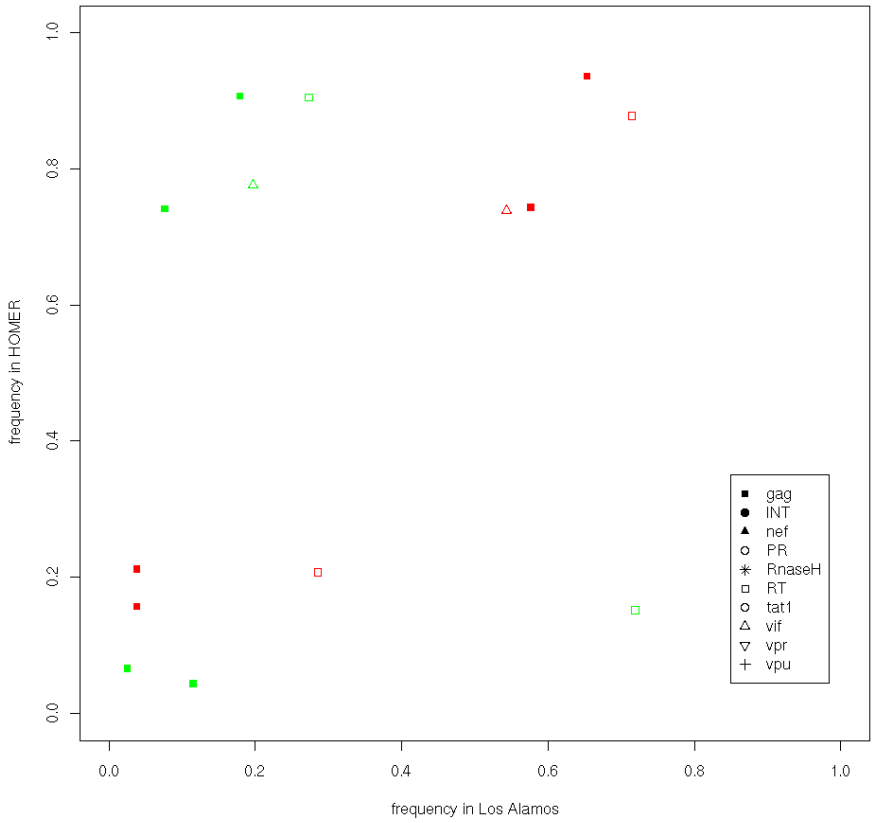


Figure 3.13: Frequencies of significant mutations within the two datasets Los Alamos and HOMER. Red points depict the frequencies in X4-viruses, green points the ones in R5-viruses. Different symbols were used to discriminate individual regions.

dataset while the frequency of this amino acid in the HOMER dataset is described by the y-axis. Only proteins for which data in both datasets was available are shown. One can see that the frequencies within X4-viruses are in good agreement between the two datasets ($r^2 = 0.914$), however the frequencies in the corresponding R5-isolates differed substantially ($r^2 = 0.019$).

#### Differences in individual parameters

In total, 64 features were significantly different between phenotypes before correction for multiple testing. 16 of these remained significant after correcting with the Holm-Bonferroni method. Among these were 6 derived from the Los Alamos dataset and 10 from the clinical dataset (see Table 3.8). A bit surprisingly, only eight of the sixteen properties were located in the envelope protein.

As already seen when analyzing specific amino acids, the most pronounced differences were found in the *gag*-protein of samples derived from the Los Alamos HIV Sequence Database. There, the number of polar residues differed by four, on average. However, as

| *region* | *feature* | *mean in R5* | *mean in X4* | *p-value* |
|---|---|---|---|---|
| gag | polar | 263.85 | 267.96 | $4.49 \cdot 10^{-5}$ |
| C2c (HOMER) | # glyc.sites | 1.71 | 1.91 | $1.00 \cdot 10^{-4}$ |
| C3 | negative | 5.53 | 5.15 | 0.0003 |
| C1a (HOMER) | tiny | 2.41 | 2.95 | 0.0006 |
| rre | tiny | 26.94 | 26.25 | 0.0008 |
| TAR | small | 11.92 | 12.89 | 0.0008 |
| C3b (HOMER) | positive | 5.92 | 6.49 | 0.0017 |
| vpr (HOMER) | length | 95.55 | 95.82 | 0.0019 |
| C2a (HOMER) | charged | 4.19 | 4.56 | 0.0024 |
| C3a (HOMER) | proline | 0.02 | 0.00 | 0.0025 |
| vpu (HOMER) | charged | 24.94 | 25.43 | 0.0026 |
| C3b (HOMER) | # glyc.sites | 2.11 | 1.78 | 0.0027 |
| C3b (HOMER) | charged | 9.67 | 10.34 | 0.0029 |
| INT | tiny | 61.60 | 60.68 | 0.0033 |
| V1 | hydrophobic | 12.30 | 13.38 | 0.0037 |
| vpu (HOMER) | negative | 14.04 | 14.48 | 0.0042 |

Table 3.8: Features significantly different between R5- and X4-samples after correction with Holm-Bonferroni.

already encountered before, results from one dataset could not be reproduced on the other dataset. No feature tested for difference in the HOMER cohort proved to be significantly different in the gag protein. Similarly, of the 16 properties found to significantly different after correction for multiple testing were 5 associated with the conserved region 3. However, while in the clonal dataset, the number of negatively charged residues was correlated with phenotype, this result could not be replicated on the clinical dataset.

### 3.3.3 Discussion

In this brief analysis, we have shown that there are amino acids associated with HIV-1 coreceptor usage in other regions than V2, V3, and gp41. We not only found residues in the envelope protein gp120 to be correlated with R5- and X4-viruses but also mutations in the polymerase, the gag gene, and the regulatory proteins. However, due to the lack of data one should not overemphasize our results but consider them with care. They may be valuable resources for more focused experiments or when more data becomes available.

The Holm-Bonferroni method used for correction for multiple testing discarded around 99% of the residues found to be significant for a phenotype. This might be a bit too conservative so that one should think about using a better method in the future. On the other hand, when testing data from the V3 loop and gp41 we found several mutations that remained significant after correction for multiple testing despite the fact that we used the even more conservative Benjamini-Hochberg method. Thus, we can assume that correction for multiple testing is not a substantial problem in this analysis.

For the V3 loop, 54 residues were associated with tropism after using Holm-Bonferroni indicating that multiple testing has not discarded too many features.

What was apparent was the high correlation of amino acids or other properties of the gag protein with tropism. Although neither of these mutations could be confirmed when using data from the other the dataset, one should not neglect them completely. Instead, because gag encodes for the structural proteins being in contact with the envelope proteins gp120 and gp41 a link between tropism defining mutations in proteins binding to each other would not be completely surprising. Therefore, more detailed analysis of this gene should be done in the future when more data becomes available.

Unfortunately, when comparing our results on the two datasets used for analysis, we could not find good agreements. Instead, while the frequencies of amino acids significantly associated with coreceptor usage were highly correlated in X4-viruses, there were major differences between the R5-isolates of the Los Alamos dataset and the HOMER cohort. This is especially regrettable because about 85% of the patients in the HOMER cohort harbored R5-viruses indicating that there is a major difference between the sequences in these two datasets (see also Section 5.4). If these differences are due to exposure to antiretroviral drugs of the isolates in the Los Alamos HIV Sequence Database or because of other reasons remained unclear. Another possible bias between the two datasets might be the different prevalences of subtypes they contain. While the Los Alamos dataset omprises a diverse set of subtypes, almost all samples from the HOMER cohort are from subtype B viruses.

The obvious differences between the two datasets refrained us from incorporating any mutation found to be significantly associated with coreceptor usage in this analysis into a prediction system. Nevertheless, we still think that it makes sense to monitor all HIV-proteins for associations with phenotype in the future because they might indicate an evolution of the viruses towards X4. With more data becoming available one might also understand and resolve the differences between the two datasets used here (see also Section 5.4). Understanding the differences might lead to a better confidence in the amino acids found to be significant in this analysis and eventually lead to their final incorporation into coreceptor usage prediction tools.

# 4 Massively Parallel Sequencing in the Realm of HIV-1 Coreceptor Usage

Massively parallel sequencing, sometimes also referred to as ultra-deep sequencing, is a method in which as its name already says lots of molecules can be sequenced in parallel. It facilitates sequencing millions of viruses from a single plasma sample affording to a more detailed picture of the viral quasispecies and its minor variants. When the technology became available, many people thought that it would solve most questions regarding evolution of HIV-1 and prediction of resistance or coreceptor tropism. Patients failing under HAART were often thought to have minor resistance mutations at the start of therapy that were previously undiscovered with standard Sanger sequencing technologies. These should be demasked with this new powerful technology such that a more effective drug regimen could be selected for these patients. In the realm of coreceptor usage prediction, the hope and expectations were even higher. This was mainly because prediction systems inferring tropism from genotype worked very well with clonal samples, but relatively poorly when tested on real clinical isolates. It was assumed that the reason for this were mainly minor populations being much more important than for HIV drug resistance. Why minor variants should be more important here than in the case of drugs targeting protease or reverse transcriptase, is of course a good question. One logical answer is that the envelope protein is much more variable both between different patients and within a single patient than the traditional drug targets derived from the pol gene.

While virologists had high hopes they had soon to encounter a major problem with this new sequencing approach: they simply generate too much data to analyze it easily. Just to give an impression about how things changed in recent years: At the beginning of this thesis we had a database of 1,100 genotype-phenotype pairs. Nowadays, with the help of ultra-deep sequencing we can generate up to 1.2 million sequences in one sequencing run from a single patient sample. It is therefore obvious that bioinformatics has become much more important and is currently one of the main bottlenecks in this field.

Within this chapter, we will first give a short introduction into the 454-technology (Section 4.1). There are two reasons for concentrating on this technology: first of all, because it is the only one we got data from, so far, and second because the 454-machine is capable of sequencing longer reads which span all of the V3 loop thus do not have to be assembled. This is more easily to handle and allows for inspecting on haplotide data directly. Following this introduction, some technological issues when dealing with ultra deep sequencing data will be discussed (Section 4.2). Thereafter, we will present how we can improve prediction of coreceptor tropism with this new technology (Section 4.3). And last but not least, we will describe the web-server we have set up to predict HIV-1 tropism from 454 sequence data (Section 4.4).

## 4.1 The Genome Sequencer FLX from 454 Life Sciences

The Genome Sequencer FLX is a next generation sequencing platform commercialized by 454 Life Sciences (Branford, CT, USA). High-throughput DNA sequencing is achieved using a novel massively parallel sequencing-by-synthesis approach.

The original GS FLX system has a throughput of 100 million base pairs and an average read length of more than 240 bases. In October 2008 , 454 released the Genome Sequencer FLX Titanium Series reagents featuring 1 million reads with an average length of 400 base pairs.

In total, 400-600 million bases can be sequenced in a 10-hour run. However, in order to reduce the high costs of an individual run, one can use molecular barcodes termed multiplex identifiers (MIDs). These are used to specifically tag samples with a unique sequence (barcode) at the start of each read. After the run, the reads can be sorted to the different samples by searching for the respective barcodes.

A 454-run is composed of four steps:

1. the generation of a single-stranded template DNA library (not applicable in amplicon sequencing)
2. the emulsion PCR
3. the sequencing-by-synthesis step, and
4. the data analysis

In the following these will be described in more detail.

**DNA library preparation** Genomic DNA is fractionated into smaller fragments of about 300- to 800-base pairs. Short adapters (A and B, fusion primers in case of amplicon sequencing) of 44 bases are then ligated onto the ends of these fragments. These adapters, specific for both the 3' and 5' ends, are composed of a 20-base PCR primer, a 20-base sequencing primer, and a 4-base key sequence which is used for read identification, signal normalization, and quality assessment. Adaptor B contains a 5'-biotin tag for immobilization onto streptavidin-coated beads.

**Emulsion-based clonal amplification** In the first step of the emulsion PCR (emPCR), the library material is mixed at limited dilution with beads carrying one of the adapter primers on its surface in a water-in-oil emulsion. Because of the limited dilution of the library material, only single DNA molecules will bind to a single bead. The DNA molecules bound to the streptavidin beads are emulsified with the amplification reagents in a water-in-oil mixture.

Each bead is then encapsulated together with PCR primers, nucleotides, and polymerase into individual droplets forming their own microreactor in which parallel DNA amplifications are performed with normal thermal cycling. Because PCR amplification of a specific fragment occurs only in its specific oil droplet, the DNA fragments are clonally amplified so that in the end several million identical copies of the single-molecule DNA template are

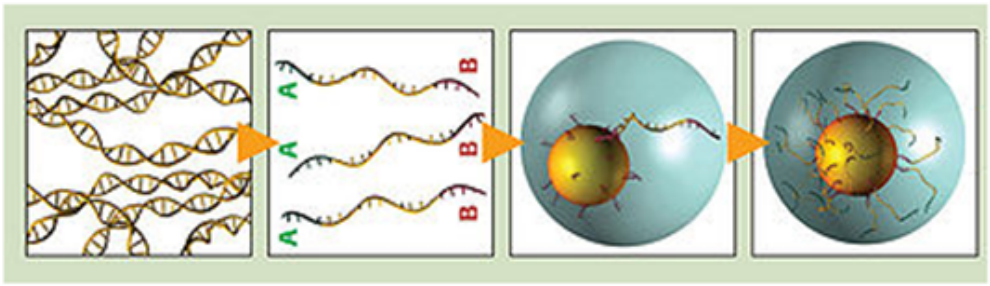


Figure 4.1: Different steps of DNA library preparation and emulsion PCR. Genomic DNA is first isolated, fragmented and ligated to adapters. Fragments are then bound to beads under conditions that favor one fragment per bead. The beads are encapsulated in the droplets of a PCR-reaction-mixture-in-oil emulsion and amplified so that each bead is carrying millions of copies of a unique DNA template. Figure derived from http://454.com/products-solutions/how-it-works/index.asp.

bound to each bead. Finally, the emulsion is broken and the template-carrying beads are enriched using magnetic streptavidin beads that select for biotin-containing beads.

**Sequencing-by-synthesis** The enriched template-carrying beads are layered onto a picotiter plate and centrifuged to deposit the beads into the wells. The wells are then overlaid with a mixture of smaller beads that carry immobilized ATP sulphurylase and luciferase necessary to generate light from free pyrophosphate. These create the individual sequencing reactors.

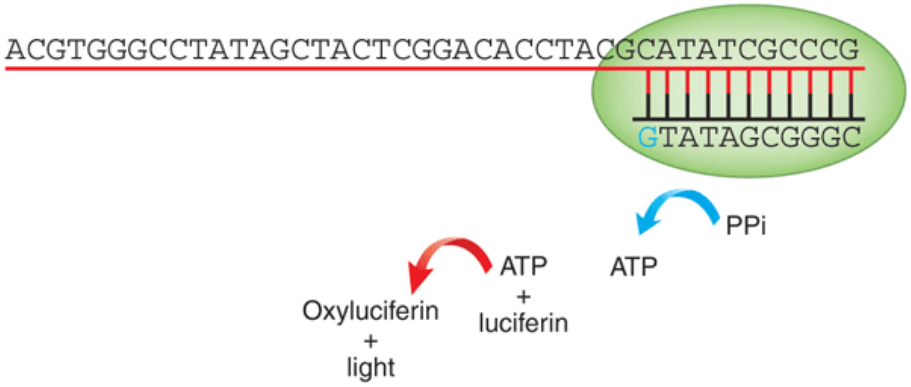


Figure 4.2: Principle of sequencing-by-synthesis. Upon incorporation of one or more nucleotides, pyrophosphate is generated ($PP_i$) and converted to ATP by sulphurylase. ATP drives the oxidation of luciferin by luciferase so that light is emitted which can then be measured with a CCD camera. Figure from (Rothberg and Leamon, 2008) with kind permission from Nature Publishing Group.

For sequencing, nucleotides are flown across the picotiter plate sequentially in a fixed order. DNA synthesis is carried out in real time in each well and positive incorporation of one or more nucleotides at a given flow of a particular nucleotide generates pyrophosphate. This is converted to ATP by sulphurylase and subsequently drives the oxidation of luciferin

by luciferase (see also Figure 4.2). As a consequence, light is emitted and captured by a charge-coupled device (CCD) sensor. The light intensity is proportional to the number of bases incorporated into the strand. After a flow of each nucleotide, a washing step is done to ensure that nucleotides do not remain in any well before the next nucleotide is introduced.

**Data analysis** After the sequencing run, the raw signals are background-subtracted, normalized and corrected. Normalization is performed using the 4-base key sequence at the start of each read. For a particular well, the normalized signal intensities indicate the number of nucleotides, if any, that were incorporated at each nucleotide flow. These can be displayed in a flowgram (see Figure 4.3).

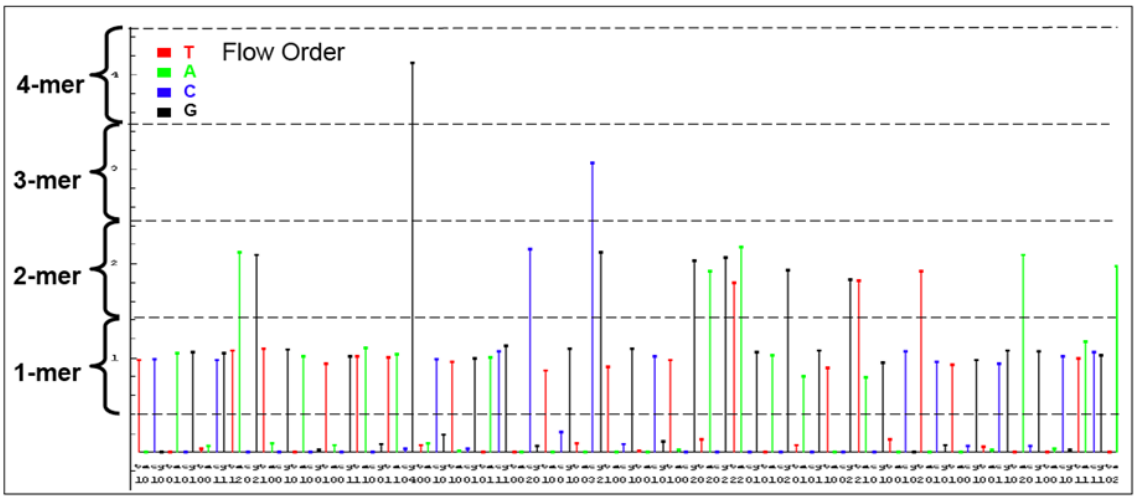


Figure 4.3: Flowgram of one read in a 454-sequencing run. Each column represents the signal intensities from a nucleotide flowed over the picotiter plate. The first four bases depict the key sequence (TCAG here), used to identify wells containing a DNA-carrying bead and for normalization. Derived from http://www.454.com/products-solutions/how-it-works/sequencing-chemistry.asp.

The light intensity is usually proportional to the number of bases incorporated into the strand. However, due to leftover nucleotides in a well or to incomplete extension, synchronism can get lost among the templates of a bead. This leads to sequencing errors predominantly appearing in homopolymer nucleotide stretches. Data is stored in standard flowgram format (SFF) files for further downstream analysis.

## 4.2 Technological Aspects when Dealing with Massively Parallel Sequencing Data

### 4.2.1 Dealing with the Amount of Data

The amount of data generated by the next generation sequencers raises several problems. First of all, storage can become an issue over time. While raw-graphic files containing gigabytes of data per run are usually not stored, this is not necessarily true for the *standard-flowgram-files* (*sff*-files) that have around 300Mb per run. To save space, these files should be further split into *fasta-like* (*fna*) files and *quality* (*qual*) files. In some cases (e.g. assemblies), the quality values are needed, however, for most analyses in the realm of this work they can simply be deleted. One can use the quality scores to better detect sequencing errors, but the bioinformatics task is usually tough enough already so that we focus on other aspects of the analysis. Although the resulting fna-files can still be around 50Mb per run, this should be a minor issue in times of cheap hard-disc space. If not, one can use compression algorithms like zip to reduce the required disc space.

The real problems begin when processing the data locally and with web-tools. There are several computational bottlenecks and problems when sending the data over the internet. The computational costs can become very impressive: suppose we have a titanium run with an average read-length of 500 base pairs and 1.2 million reads generated. We want to align all reads against a reference sequence of length 300. Using standard pairwise alignment methods, we would compute 500 times 300 matrix-entries per alignment. That would be (just for the loops) $1.2mio \cdot 1.2mio$ computations - even in times of dual-core processors a very large number.

It is relatively easy to reduce this amount of computation and probably most people will immediately find the same or similar solutions as we will describe here. There are also more and more implementations becoming available which can deal with next generation sequencing data. Unfortunately, these do not necessarily fit always to the problem at hand and at the time we started working with this kind of data only a few of them were available. Therefore, for the sake of completeness, several approaches we are working with are mentioned here. These ideas are surely not the perfect solutions and they will not be effective for every task. However we think that they are a good starting point when dealing with this kind of data.

**Combining similar variants** The simplest idea but certainly also the most often used and most effective is to combine sequences that are similar. If one has a sample in which 100% of the reads are the same, one only has to align them once to a reference sequence.

Suppose we have 12 sequences in a sample as shown in the example given in Figure 4.4. These can be collapsed to four different variants which have to be aligned and processed. Thus, the computational costs for aligning and processing the 12 sequences would be reduced by a factor of 3.

However, often we have small deviations between the different viruses even in a highly conserved population. These can be synonymous nucleotide mutations that will not have an impact for most analyses and alignments. Hence, a further step is to translate the sequences and work on the protein instead of the nucleotide level. In our example, the

```
AGTACGAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
AGGTCAAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
AGACGTAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
AGACGAAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
AGACGAAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
AGGTCAAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
AGACGTAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
AGTACGAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
AGTACGAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
AGTACGAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
AGACGTAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
AGACGAAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA
```

Figure 4.4: Example for how to reduce the amount of data to be processed.

number of variants can be reduced by combining the last two nucleotide sequences to a single amino acid variant.

```
AGTACGAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA     4 times
AGGTCAAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA     2 times
AGACGTAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA     3 times
AGACGAAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTA     3 times
```

Figure 4.5: The 12 sequences given in the example of Figure 4.4 collapsed to four different variants.

```
STSLAEEEVVIRS     4 times
RSSLAEEEVVIRS     2 times
RRSLAEEEVVIRS     6 times
```

Figure 4.6: After translation of the four different variants in Figure 4.5 one can combine them to three amino acid variants.

Finally, if one knows of highly conserved regions within the reads one is aligning, one can speed up things once more and sometimes quite drastically. The idea here is to split the sequences into fragments. E.g. in our example described above, instead of aligning the whole sequences we could split the sequences into two fragments, one comprising the first three amino acids, and the other comprising the remaining residues. The latter fragment is identical in all reads and therefore has to be aligned just once. While we cannot collapse more variants in the first region, we still saved much computational cost. This is because we now have a much smaller matrix to fill in the alignment step: 3 times the length of the reference to be aligned against.

In our example, without pre-processing we would align 12 sequences to a reference. Each alignment matrix would consist of 39 (the length of our reads) times 39 = 1521 entries (just for simplicity we suppose the same length for the reference). Thus, 12 times 1529 matrix = 18348 entries would have to be computed in total. When applying the aforementioned

```
STS            4 times
RSS            2 times
RRS            6 times
LAEEEVVIRS    12 times
```

Figure 4.7: Splitting the amino acid sequences into conserved and variable parts can reduce the computational costs once more. In the example of Figure 4.6, all variants only different in the first three residues. The remaining residues only have to be aligned once.

approaches, we could first translate the sequences and combine them to similar variants resulting in three different amino acid sequences. We would then split these into three different variants (*STS*, *RSS*, and *RRS*) of three amino acids and one variant comprising 10 residues (*LAEEEVVIRS*). Finally, we would align these four variants to the respective reference sequences, i.e. three times compute the scores in a matrix of $3 \cdot 3 = 9$ entries and once with a matrix containing $10 \cdot 10 = 100$ entries. In total, we have to compute just $3 \cdot 9 + 1 \cdot 100 = 127$ matrix entries which is more than a factor of 100 won in comparison to computing 18348 when performing no pre-processing.

Unfortunately, while this approach can be very effective, it is not always easy to implement. This is mainly due to the fact that we have to know the conserved regions and have to find them before computing the alignment. Although this sounds easy it is unfortunately often not so. However, the longer the reads are and the longer the conserved region is, the more time can be saved and the more interesting this approach becomes.

Last but not least, we might not be interested in the full sequences but only in parts of them. We then can further trim the sequences to 1) increase the probability of finding unique variants, and 2) reduce the alignment matrix.

Of note, these approaches work very well when analyzing amplicon data but they are rather inefficient when performing shotgun sequencing where the individual reads are very diverse.

**Improving alignment efficiency** The alignment used for GENO2PHENO[CORECEPTOR] is based on a sequence-to-profile alignment taking amino acid sequences as input. The reason for using amino acid instead of nucleotide sequences as well as using a sequence-to-profile alignment instead of a pairwise alignment against a reference strain is the high variability of almost all positions in the V3 loop. A good example for this problem is the V3 loop of HXB2, a strain that usually is taken as reference. In comparison to the consensus sequence of subtype B, this strain has a double insertion after position 14 and a deletion at position 26 of the loop. CLUSTALW (Thompson et al., 1994) aligns these two sequences as follows:

```
CONSENSUS B       CTRPNNNTRKSIHI--GPGRAFYTTGEIIGDIRQAHC
HXB2              CTRPNNNTRKRIRIQRGPGRAFVTIGK-IGNMRQAHC
```

Obviously, the alignment makes sense. The insertion is correct and the lysine is aligned to position 25 of the consensus sequence whereas the first isoleucine is deleted. One could also think about aligning the lysine against the first isoleucine and shifting the deletion to position 25, but this does not make sense in the realm of HIV-1 V3 loops in which position

25 is very crucial. While this position almost always displays a charged residue, position 26 is very hydrophobic. Moreover, a positively charged residue is a clear indicator of an X4-virus.

Unfortunately, if we replace consensus B with the overall consensus sequence of the V3 loops stored in the Los Alamos HIV Sequence Database, ClustalW aligns the lysine to position 26. The two references differ only marginally and substituting the glutamate at position 25 by aspartate (the predominant residue at position 25 occurring almost as often as glutamate at this position in subtype B viruses) would have yielded the same result. Based on this observation, we know that we have to be very careful when choosing a reference strain and using a pairwise sequence alignment. On the other hand, if we use a sequence profile, we can provide lots of information to the alignment procedure which then easily recognizes that a positively charged residue always (at least for V3 loops with a length of 35 residues) occurs at position 25 and not at position 26.

```
CONSENSUS-ALL     CTRPNNNTRKSIRI--GPGQAFYATGDIIGDIRQAHC
HXB2              CTRPNNNTRKRIRIQRGPGRAFVTIG-KIGNMRQAHC
```

While sequence-to-profile alignments guarantee stability they are quite slow because a query sequence has to be aligned against each sequence in the profile. Since the alignment is a bottleneck per se when dealing with data generated by massively parallel sequencing this can become really problematic. To circumvent this problem, we have adapted the Needleman-Wunsch algorithm to work with sequence profiles. We wanted to reduce the time complexity and, at the same time, maintain the advantages of a sequence-to-profile alignment. The basic idea behind this modification is relatively easy: Instead of aligning against a specific amino acid in the reference, we compute a combined score out of all amino acids occurring at this position in the profile. We then weight the individual substitutions by multiplying them with the frequencies of the respective residues at this position. E.g. if we have a position where amino acid A occurs in 75% of the sequences and amino acid B in 25%, we compute the substitution score for amino acid C as: $score_{pos}(C) = D_{C,A} * 0.75 + D_{C,B} * 0.25$. Because geno2pheno[coreceptor] aligns sequences with *ClustalW* using the Gonnet matrix, we decided to use the Gonnet matrix in this algorithm, too.
The algorithm works as follows:

**Input:**

Profile $P$ with aligned reference sequences $refSeq_i$, $i \in 1..n$, $refSeq_i = aa_1...aa_m$

Query sequence $seq = s_1...s_k$

**Initialize:**

Initialize substitution matrix $B_{aa_1,aa_2}$:

$B_{aa_1,aa_2} = gonnet\,(aa_1, aa_2) \;\; \forall aa_1, aa_2 \in amino\ acids$

$B_{aa,gap} = 0$

$B_{gap,gap} = 2$

Precompute match matrix $M_{pos}(aa)$ as follows:

$f_{pos}(aa)$ = frequency of residue $aa$ at position $pos$

$M_{pos}(aa_1) = \sum_{aa_2} (f_{pos}(aa_2) \cdot B(aa_1, aa_2))$

Initialize alignment matrix $S[i][j]$:

$S[i][j] = 0 \quad \forall\, 1 \leq i \leq (m+1),\ 2 \leq j \leq (k+1)$

$S[i][1] = -10000 \quad \forall\, 1 \leq i \leq (m+1)$

Initialize backtrace matrix $T[i][j]$:

$T[i][j] = 1 \quad \forall\, 1 \leq i \leq (m+1),\ 1 \leq j \leq (k+1)$

**Recursion:** $\forall\, 1 \leq i \leq m,\ \forall\, 1 \leq j \leq k\ ,\ do:$

$$val_1 = S[i][j] + M_i(s_j) - gapOpenPenalty \cdot f_i(gap) \tag{4.1}$$

$$\begin{aligned} val_2 = S[i][j+1] + T[i][j+1] \cdot gapOpenPenalty \cdot (1 - f_i(gap)) \\ + (1 - T[i][j+1]) \cdot gapExtensionPenalty \cdot (1 - f_i(gap)) \end{aligned} \tag{4.2}$$

$$S[i+1][j+1] = max(val_1, val_2) \tag{4.3}$$

$$T[i+1][j+1] = \begin{cases} 1 & val_1 \geq val_2 \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

**Backtrace:**

$i = m+1,\ j = argmax_j(S[m+1][j])$

while i $>$ 0 and j $>$ 0:

$$t_{i-1} = \begin{cases} s_{j-1} & \text{if T[i][j]} = 1, \\ gap & \text{otherwise} \end{cases}$$

$$j = \begin{cases} j-1 & \text{if T[i][j]} = 1, \\ j & \text{otherwise} \end{cases}$$

set $i = i - 1$

The following should be noted:

1. If a position in the query-sequence is ambiguous, i.e. the codon can be translated to several amino acids (e.g. for AAN) then we compute the average match-score $M_{pos}(codon)$ of all possible residues for this codon in equation 4.1. In fact, this is another advantage of our approach because other alignment algorithms are rarely able to handle nucleotide mixtures.

2. We do not compute a score for introducing a gap into the profile because we want to keep the length of the alignment fixed. In principle, allowing for gaps would be possible but is currently not desired in GENO2PHENO[CORECEPTOR] as the prediction is based on a fixed alignment.

3. We do not want to start with a gap neither in the sequence nor in the profile. Therefore we initialize $S[i][1]$ with $-10000$. For other alignment problems, one might think about changing this boundary condition.

### 4.2.2 Detecting Sequencing Errors

The second major issue when working with data generated by next-generation sequencers is provided by sequencing errors and their detection. This topic is much more essential than one might assume because unusual sequences originate from replication-defective viruses, sequencing errors, or uncompromised virus sequences that just look strange.

In the past, it was thought that almost all errors coming from the 454-machine are due to insertion- or deletion-problems in homopolymer regions. In this case, a measured homopolymer stretch of a specific nucleotide contained one nucleotide more or less than in the real sequence. This does not sound very complicated to detect and correct for but unfortunately during this work we had to encounter that there are also insertions arising from other sources. The question where these inserted single nucleotides come from still remains unresolved and one can only speculate. Our experiences show that if they occur in a sequencing run they seem to always happen at the same sites of the sequences. This suggests that they are related to specific sequence patterns in the context of the sample. However, so far we could not find such a motif among different samples. Another explanation could be that something happened in the environment and interfered with the sequencing run at a specific time. However, such things should happen only rarely and not in a substantial fraction of the sequencing runs, as we observed in our measurements.

While this topic itself is very interesting it goes beyond the scope of this thesis and would necessitate more data to analyze. Thus, we only propose some ideas how to detect and how to resolve such errors.

When dealing with V3 loop data, we compute a score which aims to reflect the probability that a sequence is not a V3 loop (we call it *percentile-score* according to the similar measure calculated by the WebPSSM-website). This score is calculated as follows:

1. align the sequence against the reference profile
2. go through the aligned sequence, for each position compute:

$$d_{pos} = \sum_{aa \in acids} f_{pos}\left(aa\right) \cdot D\left(aa, seq\left[pos\right]\right) \tag{4.5}$$

where $f_{pos}(aa)$ is the frequency of amino acid $aa$ at position $pos$ in the reference profile, and the $D(aa_1, aa_2)$ is the substitution score of replacing $aa_1$ by $aa_2$. We used the Gonnet matrix for this because it was also used in our alignment algorithm. We further added the following entries to the matrix:

$$D(aa_1, stop - codon) = min\left(D\left(aa_1, aa_2\right)\right),\ aa_2 \in acids \tag{4.6}$$

and

$$D(aa_1, gap) = 0 \tag{4.7}$$

The substitution score for a stop-codon was set to the lowest substitution score of that amino acid because it should not appear in the V3 loop. On the other hand, we did not set it to a highly negative number because a stop-codon might appear in an ambigue position (e.g. TAN encoding for either a stop-codon or a tyrosine). In such a case the stop-codon would affect the V3 score too much. With regards to the

substitution score of a gap we decided to use a neutral value because insertions and deletions occur relatively frequently in the V3 loop.

3. compute the V3 score:
$$V3score = \sum_{pos} d_{pos} \tag{4.8}$$

4. transform the *V3-score* to a *percentile-score*. We have computed the V3-score for a large number ($> 100.000$) V3-sequences and used the distribution of these scores to derive the probability of being V3 or not.

While working with V3-sequence data is an area in which we have a lot of experience and where we can easily built up such reference datasets, this does not have to apply to every type of data and group. Especially when dealing with a set of different regions the described strategy can become very complex. In the PRIUS-1 (Codoñer et al., submitted) study in which we analyzed pol-sequences for resistance mutations, we had to deal with data generated with 11 forward and reverse primers, respectively. Thus, we would have had to build up 22 reference datasets and profiles. This is still manageable but one can see that this way of dealing with the data is certainly not the best. We wanted to reduce the work for the researcher as much as possible.

Therefore, we aligned the reads against the corresponding fragment of the reference virus HXB2 rather than a set of different sequences. "Corresponding fragment" means in this case that the fragment starts at the same position as the respective primer and has roughly the same length. Subsequently we computed the similarity between the two fragments and discarded reads with a similarity below a certain threshold (e.g. 90% or 65%). However, dissimilarity to the reference HXB2 can be due to actual mutations as well as sequencing errors and is highly dependent on the length of the reference. If the corresponding fragment is too long the aligned read will contain a lot of gaps at the end and therefore have a lower similarity score. Hence, we did not compute the similarity against the aligned HXB2 sequence but replaced that sequence with the consensus of the aligned reads. In addition, we only took positions into account which were covered by at least 90% of the reads. This led to a normalization of the similarity scores and allowed for setting up a global similarity threshold of 90%.

The effect of this normalization can be seen in Figure 4.8. Every plot was generated from a single primer-region. The plots show the ordered similarity scores of all reads in all isolates in the analysis. Figures 4.8 A and B depict the similarities against HXB2 while Figures 4.8 C and D show the similarities against the consensus of each isolate. By inspecting a specific plot one can easily think of using the inflection point of the curvature as the threshold to be used. E.g. for forward primer 1 we would set it to about 0.65 while for forward primer 9 we would choose 0.9. One can also derive these cutoffs automatically, however different thresholds for different regions are always problematic to defend. Furthermore, we would like to have the thresholds already before the analysis and not only afterwards. In addition, there is a clear difference between the shapes of these two "curves". This comes from the different patient isolates. For primer 1 there seems to be more diversity among the isolates than for primer 9. The plot of primer 1 suggests that there are isolates where the similarity of the consensus is already below 0.8 while the consensus of other isolates is more similar

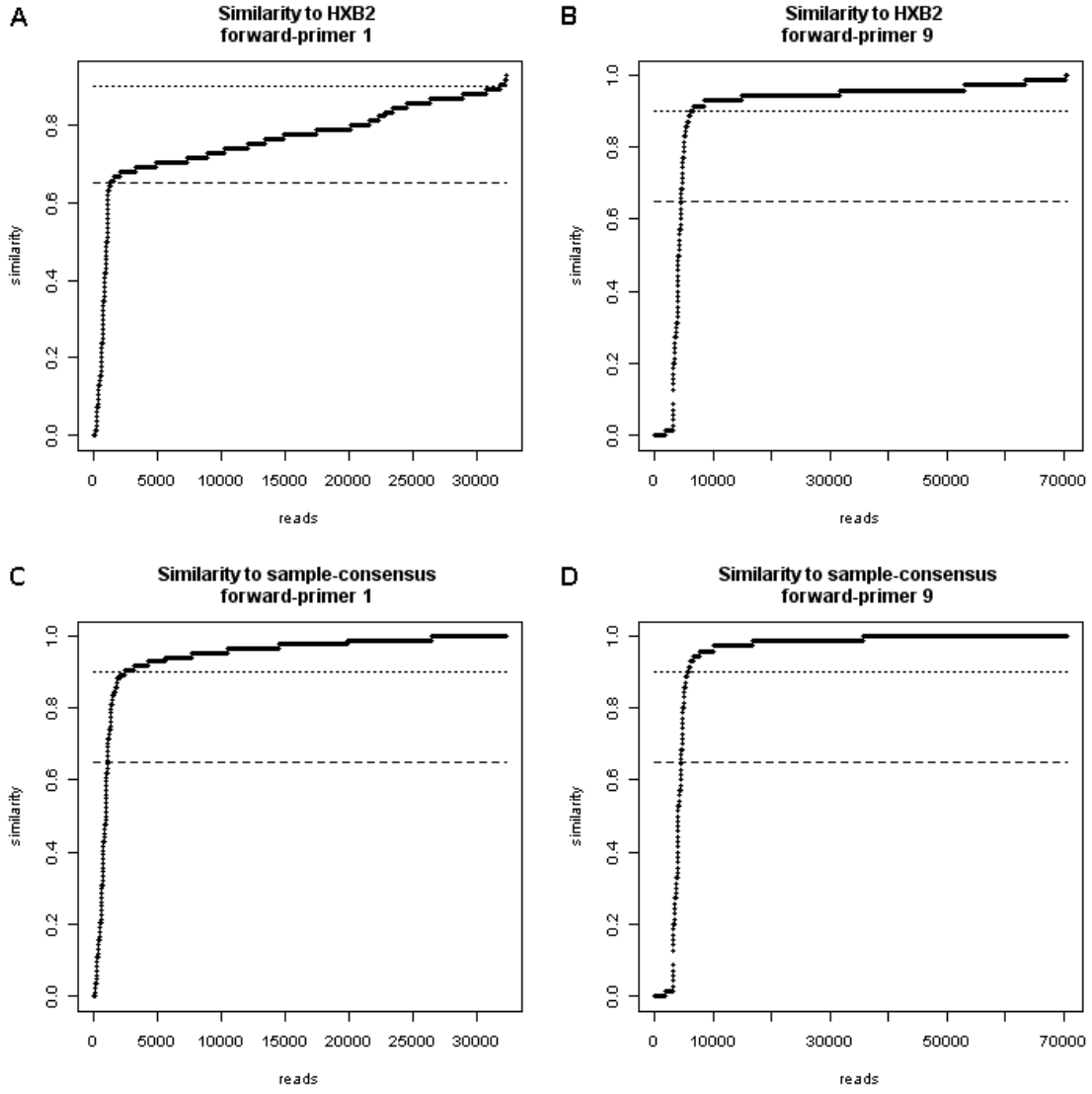


Figure 4.8: Comparison of similarity scores against HXB2 and sample-consensus from data gathered in the PRIUS-1 study. The lines indicate 65% similarity (dashed) and 90% similarity cutoffs (dotted)

to HXB2. This implies that viruses of one isolate have to be mutated much more to get filtered out than others. If we now have a look at Figures 4.8 C and D, we can see that a universal threshold of 90% is a useful threshold for all primers and directions. To improve this once more, one can of course use techniques that derive the inflection point of the curvature automatically.

### 4.2.3 Dealing with Sequencing Errors

When we have found sequences with mutual sequencing errors, how should we deal with them? Usually, we discard these reads from further analysis. Unfortunately, there are sequencing runs in which more than 50% of the reads have frame-shift errors due to single-base insertions or deletions. In such a case one might think about *fixing* these errors. While we cannot be sure that these indels arose from a problem during sequencing, we can decide safely if they are lying before or after a homopolymer stretch. Hence, if one really wants to fix and use these reads, we propose the following work flow which we used in a project dealing with HIV-sequence data covering the polymerase:

> Align the nucleotide sequences to a related reference, at best to the bulk sequence of the isolate. We propose the bulk sequence because it should be very similar to most of the reads.
>
> Go through the codons of the reference sequence and analyze their alignment. Several cases can occur, deal with them as follows:
>
> 1. There are no gaps, neither in the reference nor in the read: Do nothing and analyze the next codon
> 2. The codon covers three positions in the alignment and the query part of the alignment contains one gap: there might be a deletion due to a homopolymer stretch. Therefore, if this gap is preceded or succeeded by a homopolymer stretch, then change the gap to the respective nucleotide of the homopolymer stretch. Otherwise, change the whole codon to a new state *???*.
> 3. The codon covers three positions in the alignment and the query part of the alignment contains two gaps: change the whole codon to *???*.
> 4. The codon covers three positions in the alignment and the query part of the alignment contains three gaps: This indicates a "*normal*" codon deletion, leave it as it is.
> 5. The reference codon contains one gap: The nucleotide insertion might arise from a problem with a homopolymer stretch: If the nucleotide in the query sequence at the position of the gap is succeeded or preceded by a homopolymer stretch of the same nucleotide, then delete this nucleotide. Otherwise change the codon to *???*.
> 6. The reference codon contains two or more gaps: change the query codon to *???*.
> 7. If there are gaps between two reference codons: if their number is dividable by three, then leave it as it is, otherwise change the query part to *?*'s and add additional *?* until their number is dividable by three.

The different correction rules are shown in the example given below:

```
reference:          GCA|GGG|T-TA|-|GGG|-|CCC|---|GA--A|T-GG|GGG|CCC|TGT|TTT
query:              GC-|GGG|TCTA|G|GGG|A|CCC|CCC|GAGAG|TGGG|GGG|---|--T|TTT
rules:               2   1   5   7  1  7  1   7    6     5   1   4   3   1
```

Figure 4.9: Alignment of a read against the reference sequence before error correction. The |'s denote the codon-boundaries in the reference sequence. The described rules to be used are shown, too.

After applying the rules described before, the corrected alignment looks like this:

```
reference:          GCA|GGG|TTA|GGG|---|CCC|---|GAA|TGG|GGG|CCC|TGT|TTT
corrected query:    GCG|GGG|???|GGG|???|CCC|CCC|???|TGG|GGG|---|???|TTT
```

Figure 4.10: The alignment after error correction. *???* denote codons which are flagged suspicious and which should not be analyzed further.

Codons with triple question marks can then be either used to filter out reads or translated to an amino acid *?* indicating that this position of the read should not be further analyzed. While for coreceptor usage prediction this would make no sense because all positions of the V3 loop are used in the prediction system, in other cases one might be happy with this approach because one is not interested in all positions and can still use the other positions of the read.

Of course, this is a problematic solution and should only be used if too many reads would be discarded otherwise. The approach is error-prone and there might be several cases where it will fail. The first residues in the described example are actually for a case which really occurred in one of our works and where our approach failed. The reason is that if we had aligned the query against the reference without introducing any gaps this would have led to the same three amino acids *AGL* at the start of both reference, and query. By applying our approach on the other side, we would have "corrected" to *AG?*.

Hence, one should add additional checks to this procedure. We added the following:

1. if an insertion or deletion is changed then first check if the resulting amino acid occurs in a reference database with a specific frequency

2. before doing this analysis check if the inserted gaps are really necessary: go through the alignment, if you find a gap then search for the next position containing a gap in either the query or the reference. If both gaps are in the same sequence then take the second gap-position and repeat, otherwise: check if the removal of these two gaps would yield a query fragment where the translation can be found in a reference database. If so, then remove both gaps and search for the next gap position.

Last but not least, one has to define what a homopolymer stretch means. We used three consecutive nucleotides but four or five would also have been possible.

### 4.2.4 On the Presentation of 454-results

Massively parallel sequencing generates impressive amounts of data. While the properties of the technique are remarkable, one also has to think about how to present the results. On the one hand, we want to sequence as many viral strains as possible and want to give a detailed picture of the viral quasispecies, on the other hand we do not want to return thousands of prediction scores that only confuse people. Instead, we want to explain our results in as few numbers as possible. Obviously, the combination of these two demands can become a bit problematic. When inferring viral tropism or drug resistance from bulk genotype, we obtain a single prediction per sample. When we give the prediction methods thousands of sequences, we will of course also get back thousands of predictions. But how do we handle this? The easiest way is to look at the worst case: e.g. is any of the viruses predicted to be an X4-virus, or what is the highest resistance factor? Unfortunately, we will almost always get back the answer that at least one variant is resistant or an X4-virus but we do not know if this is *clinically relevant*. Can we infer from one single sequence that treatment with a drug will fail? Probably not. Moreover, one should also not forget that there might be sequencing errors not detected by our quality filters and that such a resistant strain could arise from an undetected sequencing error. Thus, we propose that one should at least calculate the percentage of viruses being e.g. resistant. When dealing with viral tropism predictions a second issue comes to mind: geno2pheno[coreceptor] computes a false-positive rate (FPR) displaying the probability of classifying a virus falsely as an X4-virus. Instead of giving back a binary prediction result "R5" or "X4", the system returns the FPR allowing the user to set the confidence level on his own. E.g. if one is willing to predict every tenth R5-virus falsely to be X4, one uses an FPR-cutoff of 10%. All sequences with a predicted FPR of less than 10% will then be classified to be X4 whereas sequences with FPRs above 10% will be regarded to be R5. The problem is that different groups use different thresholds to separate R5- from X4-viruses. Even within a specific lab, different cutoffs are used. This is because e.g. if a patient has no other drug options, the clinician will perhaps take a higher risk than for a patient who is under a successful drug regimen but who wants to change the drug due to severe side effects. Based on these thoughts, we compute the percentages of predicted X4-viruses at different cutoffs (see Figure 4.11).

The drawback of this approach is that very similar isolates can look totally different. Suppose we have an isolate sequenced for which geno2pheno[coreceptor] predicts almost all strains to be X4 with an FPR of 10.1% and a second sample from a different patient for which almost all sequences were predicted to be X4 with an FPR of 9.9%. If we used the default 10%-FPR cutoff of geno2pheno[coreceptor], we would classify all sequences with a predicted FPR of 10.1% to be R5 and the strains with an FPR of 9.9% to be X4. Thus, in case of the first patient almost all sequences would be predicted to be R5 whereas in case of the second patient we would have almost only X4 predictions. Based on these results, we would refrain the second patient from a therapy containing a coreceptor antagonist, but not the other. This might be right and at a certain point we will have to make a decision where to set a cutoff, however, we would like to give the virologist a better feeling of what the viral quasispecies looks like.

As we said before, we want to describe the viral population in as detailed fashion as

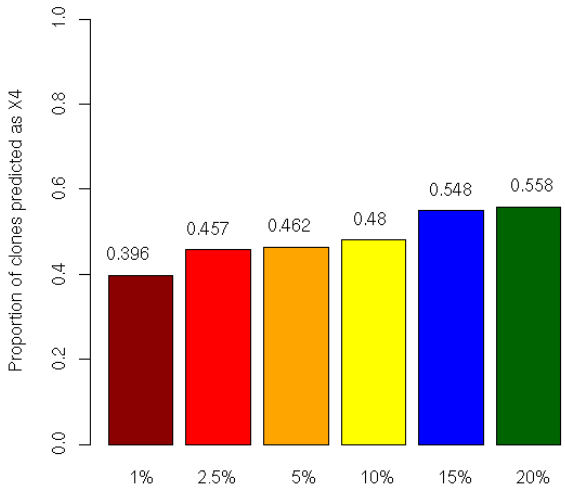


Figure 4.11: Example for a prediction report from a 454-sequencing run. The bars show the frequencies of viruses predicted to be X4 at different FPR-cutoffs of geno2pheno[coreceptor].

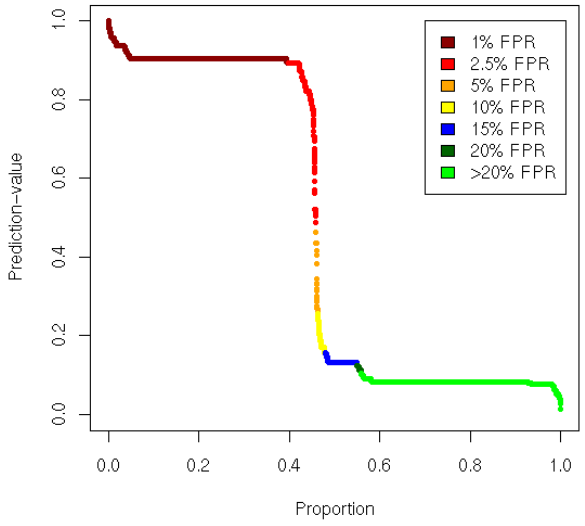


Figure 4.12: Example for a prediction report from a 454-sequencing run. Each point in the chart displays the prediction score for one individual read. The higher the score, the more probable the sequence encodes for an X4-virus. The patient samples is the same as in Figure 4.11.

possible. Thus, we computed the amount of predicted X4-viruses for all possible FPR-cutoffs as follows: we sorted the computed prediction scores and displayed them in a plot

(see Figure 4.12). One can now easily set his own cutoff and infer how many viruses are predicted to be X4. Moreover, one gets a feeling how the viral population is built up: e.g. do we have one major strain or two?

### Population Dynamics

While the described results are quite intuitive and helpful for making treatment decisions on whether to administer a coreceptor antagonist or not, they lack some information one is sometimes also interested in: do the populations of two isolates (e.g. from different time points) differ or not (e.g. has the population evolved between two time points). We have tried to solve this problem by generating a landscape of the viral population.

On the x-axis we show the predicted FPR, while the y-axis shows how similar the sequence is with regard to a reference dataset. For this purpose we use the alignment score described above. If two sequences are very different, these scores are usually also very different, whereas quite similar scores usually arise from similar sequences. The z-axis of the plot describes how frequent a specific variant appeared in the viral population.

The landscape is smoothed as follows:

$$z(x,y) = \frac{1}{N} \cdot \sum_v N_v \cdot e^{-\frac{1}{2}(x - fprScore_v)} \cdot e^{-\frac{1}{2}(y - alignmentScore_v)} \tag{4.9}$$

where $N_v$ is the number of reads of a given variant $v$ and $N = \sum_v N_v$ the total number of reads of this sample, $fprScore_v$ refers to as the predicted FPR of variant $v$, whereas $alignmentScore_v$ is the corresponding alignment-score of this V3 loop.

Figure 4.13 depicts such a representation of a viral quasispecies. One can easily identify two major strains that are quite similar at the sequence level (V3 alignment score of about 145) but with different FPRs (14.7% vs. 38.1%). Furthermore, two minor strains can be seen, one in between the two major strains and one with a lower FPR of about 6%.

We can use these plots to get an overview of the viral population. When we want to compare two isolates we can simply incorporate the two plots into an animation or compute the difference between them (see Figure 4.14). With the latter we can see the shifts in the viral population from one isolate to the other quite easily. In the given example, the different variants of Figure 4.13 have all vanished (deep blue and negative z-values) and been replaced by a new major strain (red and positive z-values) that is very likely an X4-virus.

### Electropherograms Revisited

In addition to the reports on coreceptor usage prediction, we have also tried to compare the output of the 454-machine with the electropherograms generated by conventional Sanger sequencing. There were two main reasons for doing this: First, we wanted to check how reliable the generated data really is. As mentioned before, the GS FLX system generates sequencing errors, mainly in the neighborhood of homopolymer stretches. If specific unusual nucleotide mutations occur in most reads of a sequencing run, one could check them by comparing with the bulk sequence. Another reason was that we wanted to get an impression how sensitive Sanger sequencing really is. Often, one refers to it as having a sensitivity of 15-20% but it is not clear how reliable this assumption is. Furthermore,

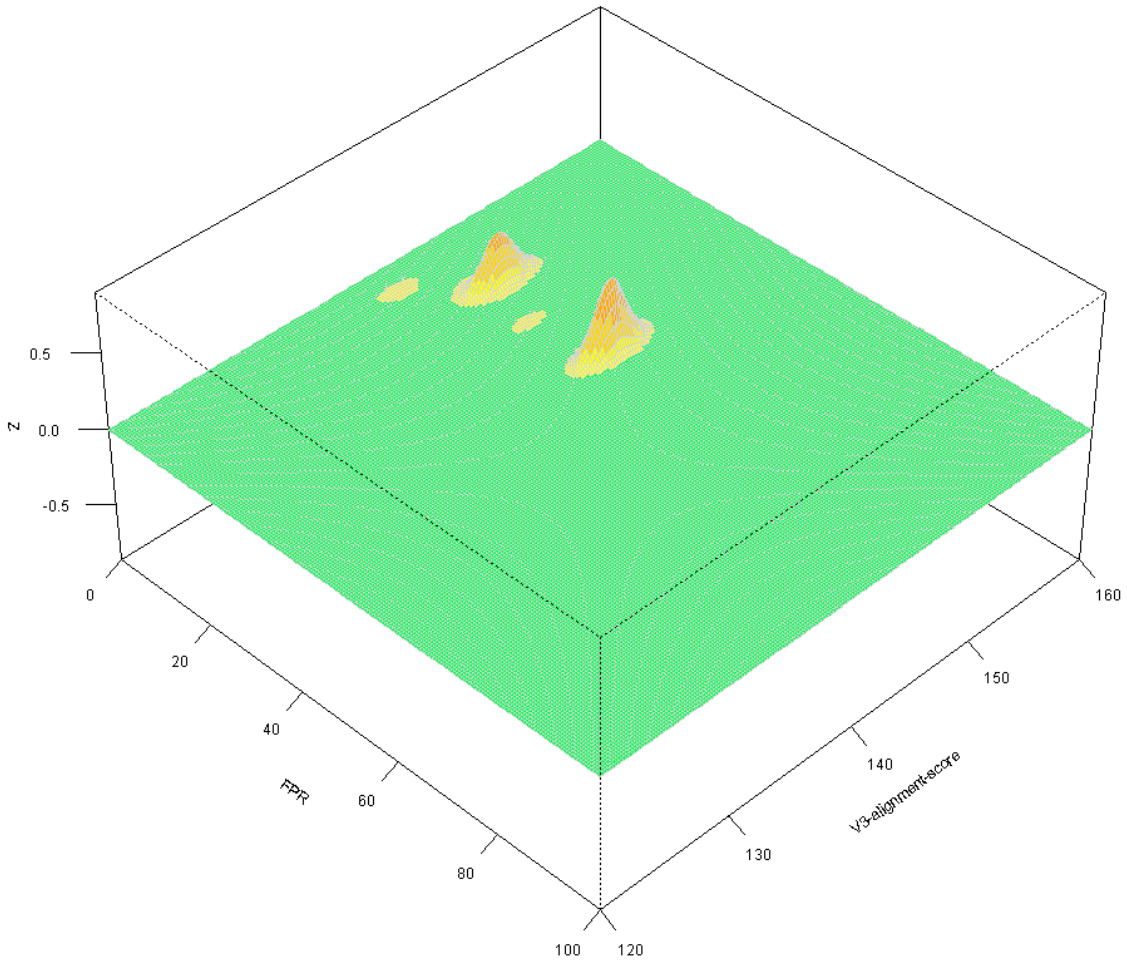


Figure 4.13: Representation of a viral quasispecies. The x-axis displays the alignment score of a virus, the y-axis the predicted false-positive rate while the z-axis describes how prevalent these viruses were in the population.

there could be different sensitivity levels dependent on the diversity of the viral population. E.g. if 4 different viruses have a prevalence of 5% each and carry the same mutation X, do we see this mutation in an electropherogram or do we only see it if a single strain has a prevalence of 15-20%?

The way we transformed the 454-data into an "electropherogram" was straightforward. For every nucleotide (A,C,G,T) we counted how often it appeared at the different positions within all reads that were subjected to be of V3.

$$N_{pos}(nt) \quad \forall pos, nt \in (A, C, G, T)$$

We then used a Gaussian function to form

$$f(x, nt) = \sum_{pos} \frac{N_{pos}(nt)}{\sqrt{2 \cdot \pi}} \cdot e^{\frac{(x-pos)^2}{2 \cdot \sigma^2}}$$

with $\sigma = 0.25$. Finally we can plot this and compare it visually with a real electropherogram (see Figures 4.15 and 4.16).

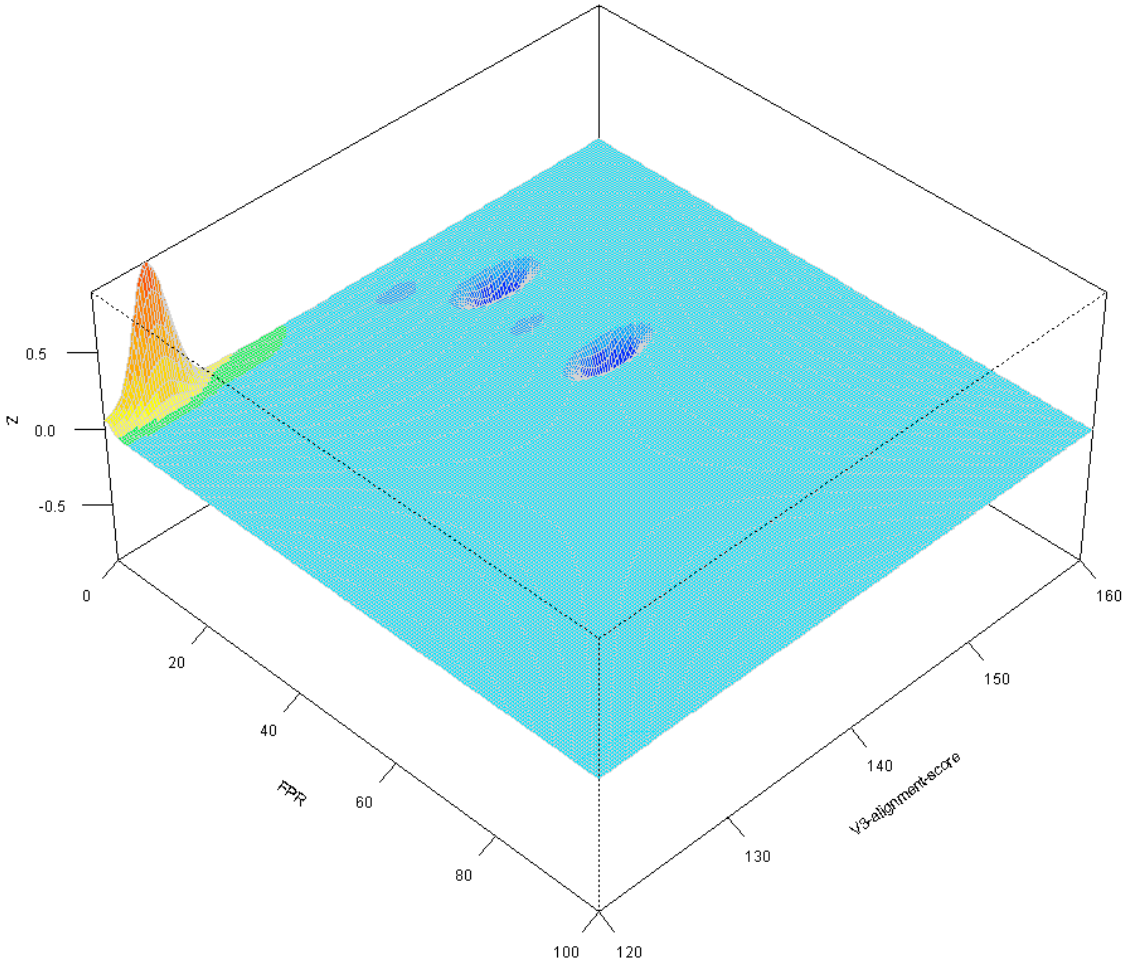


Figure 4.14: Evolution of the viral quasispecies from one time point to another.

When comparing the two plots one can see the similarities quite easily. E.g. at amino acid positions 8 and 9, both representations display nucleotide mixtures. The same holds for the middle part of the loop and amino acid position 34 in which a G and a T occur in about 50% of the samples, respectively.

The main difference which can be found is that in Figure 4.16, the peaks always sum up to the same height whereas they vary in Figure 4.15. However, it is known that the absolute peak heights in electropherograms vary and that they are context-dependent. We therefore further wanted to compare the relative peak heights with each other in order to see if the quantities of detected minor variants are similar. At every position we summed up the peak heights of the four nucleotides and calculated which proportion of the total peak height each nucleotide reached. Figure 4.17 shows a comparison of the same sample. Unfortunately, we do not have enough data for a more detailed analysis, yet. However, preliminary analysis of this sample and 10 others suggest that there is a substantial correlation between relative peak heights in Sanger sequencing traces and the quantity with which a nucleotide is detected in a 454 run.

The results of such a more detailed analysis might also be useful for developing more sophisticated sequence trace read-out approaches for population-based sequencing. Sup-

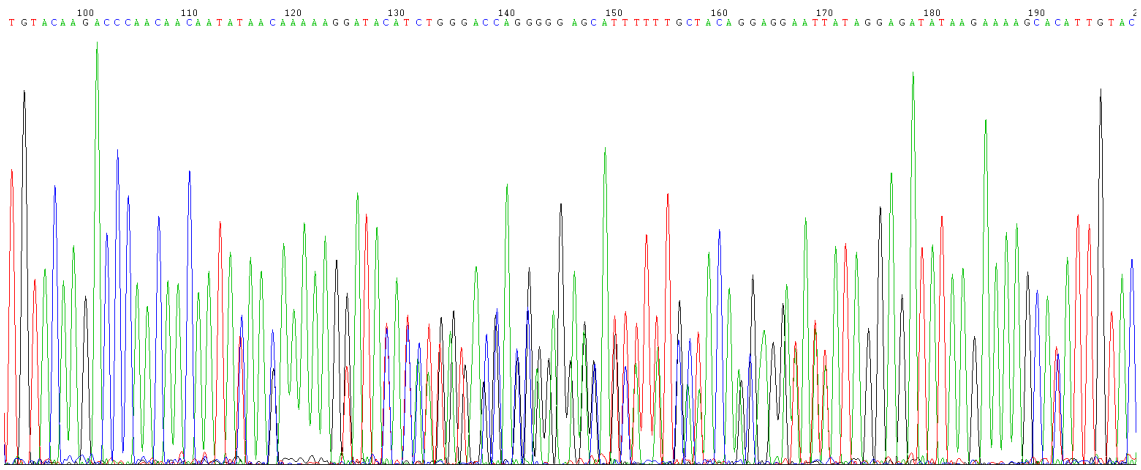

Figure 4.15: Electropherogram of the patient sample used also in Figure 4.11. The different curves denote the different nucleotides: green = A, blue = C, black = G, red = T.

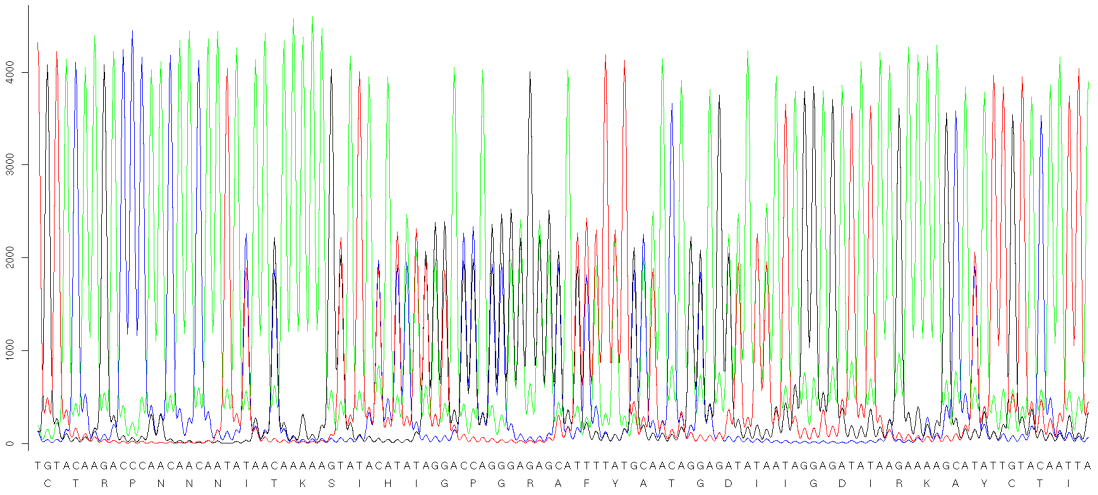


Figure 4.16: Generated "electropherogram" from the 454-reads used also in Figure 4.11. Only the V3 loop is shown, the sequence displayed below is just one possible variant. The different curves mark the different nucleotides: green = A, blue = C, black = G, red = T.

pose we have a codon with mixed nucleotides. At each position we have a high peak (e.g. a G) covering about 80% of the total size and a minor peak (e.g. an A) with about 20% of the total peak height. Currently, we get only the information that there is a mixture G/A at each positon in the codon and therefore translate this codon into all combinations from these nucleotides: AAA, AAG, AGA, AGG, GAA, GAG, GGA, GGG. This yields the amino acids glycine, lysine, glutamic acid, and arginine. On the other hand, if we could show that the relative peak heights are really displaying how frequently the different nucleotides appear in the viral quasispecies, we might argue that we have two viral variants in our example. One variant displaying a GGG codon at this position which encodes for a glycine, whereas the other variant would consist of a AAA codon encoding for a lysine. The

two additional residues (glutamic acid and arginine) we are dealing with at the moment might affect the prediction of coreceptor usage negatively so that the information of the relative peak heights would help us with the interpretation of the sequences.

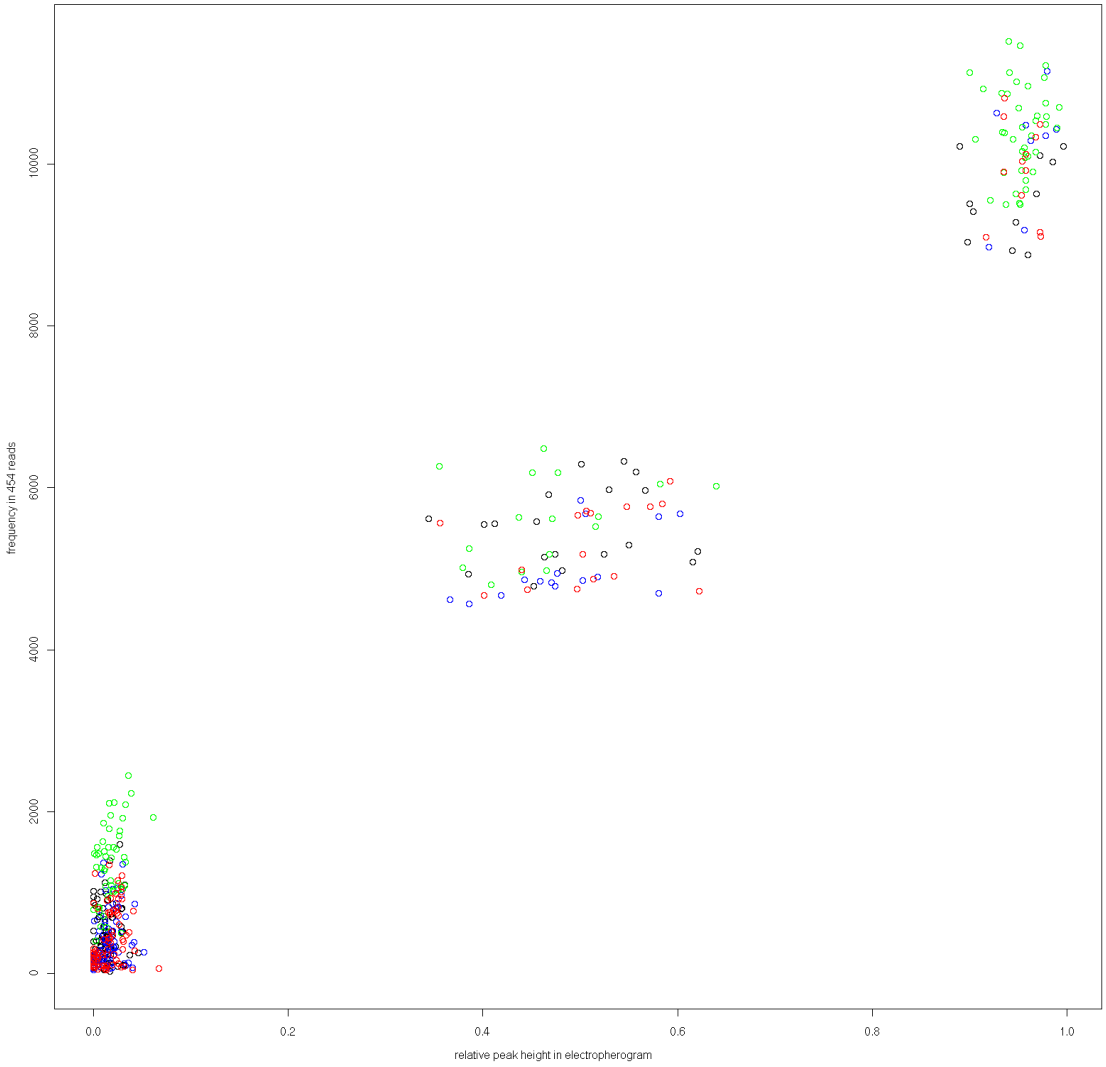


Figure 4.17: Relative peak heights of individual nucleotides in an electropherogram plotted against the frequency of these nucleotides in the corresponding 454-reads. The different colors mark the different nucleotides: green = A, blue = C, black = G, red = T.

## 4.3 Inferring Viral Tropism with Ultra-deep Sequencing

In the previous chapter we have demonstrated how prediction of HIV-1 coreceptor usage can be improved by extending the region of interest (see Chapter 3). We have shown that coreceptor defining mutations are not entirely located inside the V3 loop but also in other parts of the viral envelope gene and perhaps also in other proteins. However, we still had to encounter that the performances of genotypic prediction methods were worse on clinically derived isolates when compared to clonal data from the Los Alamos HIV Sequence Database. To some extent this is certainly due to unknown mutations in different parts of the viral envelope, however, a major factor is probably also the use of conventional population sequencing methods.

Standard bulk sequencing methods have two major problems: First, they generate a consensus sequence of the viral quasispecies. These sequences can contain nucleotide mixtures in case of virus variants differing at specific positions. From the sequence one cannot necessarily infer how the different variants looked. Suppose a sequence contains two positions with nucleotide mixtures. Each of these can be translated into two possible amino acids. From this information one cannot differentiate if two, three, or four different viruses were present in the isolate. As a consequence, the interpretation of this kind of data is more problematic than that of clonal data lacking nucleotide mixtures.

The second problem is the inability of standard bulk sequencing methods to determine minor populations. They can only detect minorities down to about 15-20% of the viral population. In contrast, phenotypic assays can reliably detect minorities of less than 5% (e.g. the Monogram Trofile Tropism Assay, Whitcomb et al., 2007). Thus, minorities of X4-viruses detected by the phenotypic approach can remain undetected when using population sequencing methods.

In comparison to standard bulk sequencing, massively parallel sequencing based on the 454 technology can resolve the viral quasispecies to almost arbitrary detail. In addition to the high sensitivity of the method, it also generates clones rather than a consensus genotype. Thus, prediction of coreceptor usage based on 454 data should be significantly more accurate than on bulk data.

In this section, we therefore performed a proof-of-concept study in which we addressed the question whether massively parallel sequencing can be successfully combined with bioinformatic approaches in order to afford improved qualitative and a quantitative prediction of coreceptor usage from the V3 loop.

The study described in this section was a joint work with Martin Däumer, Rolf Kaiser, Rolf Klein, Thomas Lengauer, and Bernhard Thiele and has been presented at the XVII International HIV Drug Resistance Workshop: Basic Principles and Clinical Implications 2008, Sitges, Spain (Däumer et al., 2008).

### 4.3.1 Materials & Methods

**Patient samples** We collected 55 samples from heavily pre-treated patients with limited therapy options screened for potential administration of maraviroc at the Institute of Virology of the University of Cologne and the Institute of Immunology and Genetics in Kaiserslautern. $3ml$ plasma from each patient was shipped on dry ice for phenotypic tropism

testing (Trofile Tropism Assay, Monogram Biosciences, South San Francisco). Plasma viral load was determined using the M2000 system (Abbott Molecular). Results were documented as either CCR5-tropic (R5), CXCR4-tropic (X4) or dual-/mixed-tropic (D/M).

**RNA extraction, cDNA synthesis and V3-PCR** RNA extraction was performed by using the Viral RNA mini kit (Qiagen, Hilden, Germany) according to the manufacturers instructions. cDNA-synthesis was performed using $10\mu l$ of RNA, specific primers V3-1 (TACAATGTACACATGGAATT, positions 6958-6977 in HXB2), V3-2 (ATTACAGTAGAAAAATTCCCC, positions 7362-7382 in HXB2) (Simmonds et al., 1990) and Superscript II (Invitrogen) according to the manufacturer's instructions in a total volume of $20\mu l$.

Amplification of the V3 region was carried out using the FastStart HiFi PCR system, (Roche, Mannheim) and primers V3-for (TGGCAGTCTAGCAGAAGAAG, positions 7010-7029 in HXB2) and V3-rev (CTGGGTCCCCTCCTGAGG, positions 7315-7332 in HXB2) (Simmonds et al., 1990) in a Primus 96 plus thermal cycler (MWG-Biotech).

Products to be massively parallel sequenced were generated with fusion primers V3FusA: GCCTCCCTCGCGCCATCAG-V3-for and V3FusB: GCCTTGCCAGCCCGCTCAG-V3-rev.

**Standard DNA sequencing and massively parallel pyrosequencing** Clean-up of the PCR products prior to standard sequencing was performed by incubation with FastAPTM in conjunction with Exonuclease I (Fermentas, Burlington, Canada) for 10 min at 37°C followed by heat inactivation for 5 min at 75°C.

Population-based sequencing of the V3 region was carried out using the Big Dye® Terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) with primers (3pmol) V3-for and V3-rev. Extension products were purified using the Biomek NXp automated sequencing reaction cleanup system (Beckman Coiulter, Fullerton, CA, USA) and were run on an ABI 3130 capillary sequencer (Applied Biosystems, Foster City, CA, USA).

Out of the 55 samples, 14 (7 R5, 7 X4) were randomly selected for further analysis with massively parallel sequencing. All 14 amplicons were purified with AMPure magnetic beads (Agencourt, Beckman Coulter) and quantified using an Agilent 2100 Bioanalyser (Agilent Life Sciences). Since bidirectional sequencing using primers A and B might result in unbalanced read composition, unidirectional sequencing with primer A (kit II; Roche-454 Life Sciences) was chosen to ensure precise quantification. After beads recovery and enrichment, approximately forty thousand beads were loaded on each region of a GS FLX PicoTiter plate subdivided with a 16-lane gasket. Sequencing was performed on a Genome Sequencer FLX (Roche-454 Life Sciences).

**Determination of coreceptor phenotype** The Trofile assay is a single-cycle recombinant virus assay developed by Monogram Biosciences. A 2.5kb long part of the patient-derived env gene is amplified by PCR and inserted into an envelope expression vector. Together with a replication-defective retroviral vector carrying a luciferase reporter gene, this vector is used to co-transfect human embryonic kidney 293 cells (HEK293). Pseudo-viruses produced by these cells are then given to engineered U87 target cell lines expressing either CCR5 or CXCR4. Upon successful infection of these cells, the reporter gene is expressed

and a light signal emitted. This can be quantitated in relative light units (RLU) and is additionally controlled by the presence of coreceptor antagonists. The standard tropism assay is reported to be reproducible and effective at detecting minority populations of CXCR4-using virus at levels as low as 5% (Whitcomb et al., 2007).

**Preprocessing of 454-data** Massively parallel sequencing data was processed directly from the Standard Flowgram Files (.sff) containing the sequence trace data. The system's methods for post-processing (GS De Novo Assembler, GS Reference Mapper, GS Amplicon Variant Analyzer) have not been used due to previous problems with this software

Two main filtering steps were applied to minimize possible errors and to sort out non-V3 sequences arising from PCR- or sequencing errors. These worked as follows: First, sequences were translated into all three reading frames. Each frame was then screened for typical start and end-motifs of the V3 loop (e.g. CTR, CIR, AHC, AYC). Only sequences where both ends of the loop were recognized in the same reading frame were further processed. The potential V3 loop was cut out and its length analyzed. While the usual V3 loop has a length of about 35 amino acids, loops containing a number of insertions and deletions have been observed previously, as well (Sing et al., 2007). To allow for such indels but limit the number of deviations from the standard loop which could occur due to sequencing problems, we introduced the criteria that V3 loops had to have a length of at least 30 and at most 45 residues. If several V3-start- or V3-end motifs were found in a specific read, only the one with the length closest to 35 was selected for coreceptor usage prediction.

**Tropism prediction from genotype** Genotypes resulting from both sequencing methods were input to a development version of geno2pheno[coreceptor] for coreceptor usage prediction (version from June 2009) (Lengauer et al., 2007). This version produces exactly the same predictions as the public version but differs in two respects: First, it can process a number of sequences in batch allowing to handle the vast amount of data generated with the Genome Sequencer FLX system and second it returns the internal scores of the prediction method which are currently not displayed on the website (the displayed FPR-scores are a transformation of these scores).

Prediction results were used in two fashions: On the one hand, cutoffs corresponding to the false-positive rates available on the website were used in order to classify the samples into R5- and X4-viruses. On the other hand, the raw values were used for detailed quantitative analysis as described below.

**Performance assessment** For performance comparison, the Monogram Trofile assay was used as reference. Performance was measured in terms of accuracy, true-positive rate and false-positive rate. The true-positive rate is the number of D/M-viruses correctly predicted as D/M while the false-positive rate is the number of R5-viruses falsely classified as D/M. Accuracy represents the total agreement between genotype and phenotype. For all evaluations, the statistical programming language R (Team, 2005) was used.

### 4.3.2 Results

**Bulk-sequencing results** Prediction results based on population-sequenced isolates were compared with tropism calls generated by the Trofile assay (see Table 4.1). Using the default false-positive rate of geno2pheno[coreceptor] (10%), 43 of the 55 isolates were predicted concordantly with the Monogram Trofile Tropism Assay. For 9 out of 22 samples phenotyped to be X4 by the Trofile assay, geno2pheno[coreceptor] predicted an R5-virus. On the other hand, for 3 of the 33 viruses determined by the Trofile assay as R5, geno2pheno[coreceptor] returned an X4-prediction. Hence, the sensitivity of predicting CXCR4-using isolates was 59.1% while specificity was at 90.9%. These prediction results are in line with results reported in recent publications (Low et al., 2008; Sing et al., 2007; de Mendoza et al., 2008; Garrido et al., 2008a; Skrabal et al., 2007). The overall concordance between the two approaches was 78.2%.

| | | geno2pheno[coreceptor] | |
|---|---|---|---|
| | | R5 | X4 |
| Trofile | R5 | 30 | 3 |
| | X4 | 9 | 13 |

Table 4.1: Comparison of genotype and phenotype. Comparison of Trofile results and geno2pheno[coreceptor] predictions from bulk-sequenced isolates tested for maraviroc administration.

#### 454-results

**General statistics** Table 4.2 shows characteristics of the individual samples analyzed with the GS FLX system. On average, around 7600 (range: 885-12992) reads were generated per isolate. The mean viral load was at 35810 copies per milliliter. There was no significant correlation between the number of reads and the viral load ($r^2 = 0.0417$). After translation and adjusting for the V3 loop we found 280 different amino acid variants of the V3 loop on average (range: 57-584). The squared correlation coefficient between the number of different V3-variants and the number of reads was $r^2 = 0.584$. When comparing R5- and D/M isolates, we did not find any significant difference between the two groups neither for the number of reads generated ($p = 0.29$, 8821 vs. 6341), nor for the viral load ($p = 0.26$, 19515 vs. 52106) or the number of found V3 loops ($p = 0.29$, 6808 vs. 4687) and their variants ($p = 0.97$, 281.7 vs. 279.1).

**Quantitative analysis of coreceptor usage** All reads in which a consistent V3 loop sequence was found were input to the geno2pheno[coreceptor] prediction system. This resulted in around 7600 coreceptor predictions per isolate. Since this vast amount of data cannot be interpreted easily we decided to display the results as described in Section 4.2.4. In addition to tabulating the numbers we also plotted the results. Figure 4.18 depicts the results for two of the isolates. Each point in these plots represents the prediction score of one read. The prediction score which can be regarded as an affinity value is plotted along the y-axis. High scores represent predictions for which geno2pheno[coreceptor] was

| *isolate* | *Trofile* | *viral load, copies/ml* | *#reads* | *#reads with V3 loop* | *#variants* |
|---|---|---|---|---|---|
| 1 | D/M | 65815 | 885 | 667 | 67 |
| 2 | D/M | 3236 | 4614 | 3205 | 233 |
| 3 | D/M | 28850 | 3875 | 2825 | 210 |
| 4 | D/M | 199005 | 12063 | 10091 | 325 |
| 5 | D/M | 10082 | 12013 | 9608 | 584 |
| 6 | D/M | 9400 | 3094 | 1581 | 147 |
| 7 | D/M | 48354 | 7847 | 4833 | 388 |
| 8 | R5 | 48000 | 10622 | 8621 | 354 |
| 9 | R5 | 9300 | 1601 | 344 | 57 |
| 10 | R5 | 16400 | 6105 | 4905 | 216 |
| 11 | R5 | 11928 | 11142 | 8677 | 301 |
| 12 | R5 | 1270 | 11551 | 8748 | 283 |
| 13 | R5 | 18709 | 12992 | 10555 | 320 |
| 14 | R5 | 31000 | 7740 | 5808 | 441 |

Table 4.2: Properties of analyzed patient isolates. Characteristics of the isolates sequenced with the 454-machine. The number of variants is the number of different amino acid sequences found for the V3 loop.

very sure of the virus being X4 (similar to a low false-positive rate) whereas low scores represent viruses probably being R5. The shape of the curve qualitatively reflects the coreceptor usage of the whole population within the relevant isolate. E.g. the population depicted in Figure 4.18 a) consists of almost exclusively R5-viruses whereas Figure 4.18 b) depicts a virus population about half of which can use the CXCR4-receptor.

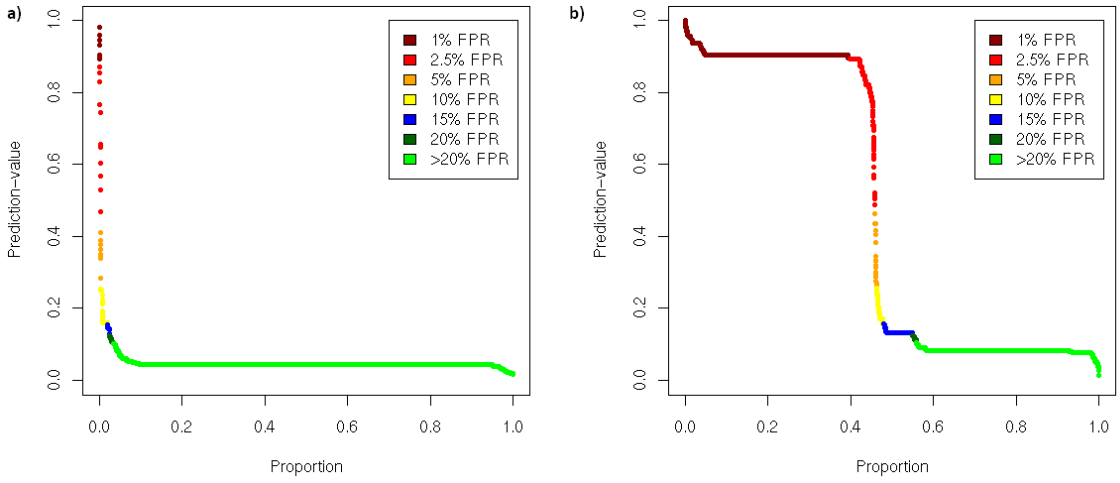


Figure 4.18: Overview of coreceptor usage predictions from one sample (a: isolate 13, b: isolate 5). Every dot represents the prediction for one read. The colors indicate at which FPR-cutoff, the read would have been predicted as X4.

In addition, we generated histograms displaying the frequency of predicted X4-viruses

for the specific false-positive rates on the GENO2PHENO[CORECEPTOR] website (see Figure 4.19 for the respective plots of 4.18). These might be useful when one is interested in the question how many viruses are predicted to be X4 at a specific cutoff. While being more easily to interpret than the previous type of graph, the global picture of the viral population is lost. E.g. from Figure 4.18 one can easily infer that the viral population is split into two major strains while this is not possible from Figure 4.19.

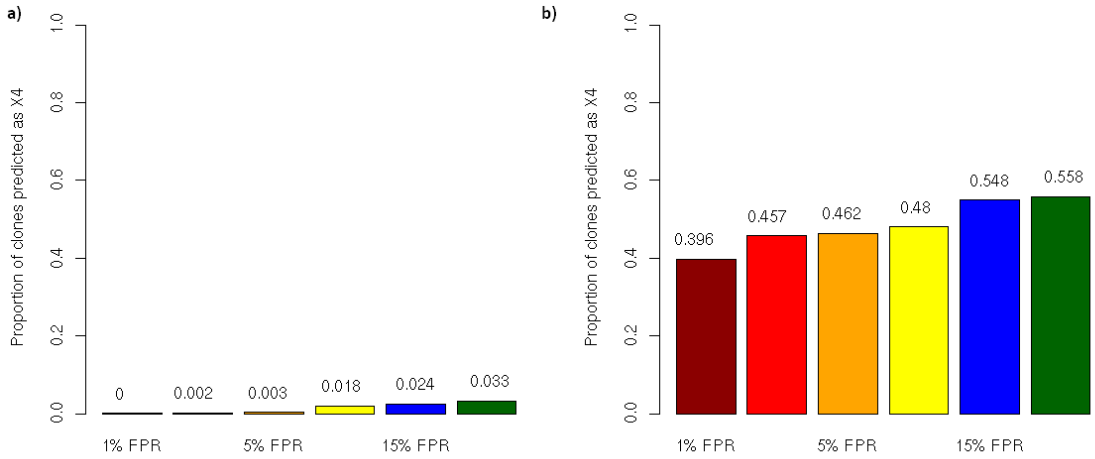


Figure 4.19: Proportion of predicted X4-viruses at different FPR in ultra deep sequenced V3 amplicons (a: isolate 13, b: isolate 5).

**Comparison between genotypic prediction and phenotypic results** In order to obtain a qualitative prediction from the ultra-deep data, two cutoffs have to be defined: First, a cutoff for the prediction score has to be selected - reflecting the affinity of the virus to the CXCR4-coreceptor. For this purpose and in order to be comparable to the bulk-sequenced isolates, we chose the cutoff resulting in the standard 10% false-positive rate of GENO2PHENO[CORECEPTOR]. Second, a minority threshold has to be set because in every sample individual clones very predictive for CXCR4-usage had been found. Three reasons for this observation suggest themselves: (i) that not every minority is clinically relevant, (ii) that the phenotypic assay also has a sensitivity limit which has to be taken into account and which is most probably higher than the limit for massively parallel sequencing, and (iii) that in some cases the 454-machine also produces erroneous sequences that are not recognized during preprocessing and prediction and which should not have an impact on the overall prediction outcome. Hence, we decided to select a minority cutoff of 5%, i.e. only samples where more than 5% of the sequences were predicted to be CXCR4-using were assigned to be X4. This value is identical with the proposed sensitivity limit of the standard Trofile assay used in this analysis. By applying these settings to the dataset, we could achieve concordance with the phenotypic assay in all but one case (isolate 7, see Table 4.3).

| *isolate* | *Trofile* | *bulk-prediction* | *MPS (% X4)* |
|---|---|---|---|
| 1 | D/M | 2.5% | 85.3% |
| 2 | D/M | 5% | 87.0% |
| 3 | D/M | 10% | 14.2% |
| 4 | D/M | 15% | 5.4% |
| 5 | D/M | 15% | 48.0% |
| 6 | D/M | — | 68.5% |
| 7 | D/M | — | 1.0% |
| 8 | R5 | 10% | 1.4% |
| 9 | R5 | — | 3.4% |
| 10 | R5 | — | 0.3% |
| 11 | R5 | — | 0.5% |
| 12 | R5 | — | 1.6% |
| 13 | R5 | — | 1.8% |
| 14 | R5 | — | 2.4% |

Table 4.3: Tropism results from the different approaches used: phenotypic assay (Trofile), coreceptor usage prediction from bulk-genotype, and coreceptor usage prediction from massively parallel sequencing. Bulk-prediction: The displayed FPR-value is the highest cutoff at which an isolate would have been predicted as X4. '—' if predicted FPR was above the highest cutoff tested (20%). MPS-prediction: a 10% FPR-cutoff was used, the value marks the frequency of X4-predictions among reads with a V3 loop.

### 4.3.3 Discussion

In this section, we analyzed the usefulness of next-generation sequencing technologies in the realm of coreceptor usage prediction. Our results show that the approach affords precise classification as well as the possibility of quantitatively assessing the tropism of the viral quasispecies. With respect to the classification, we could show concordant results with the phenotype in all but one case. It should be noted that the selected cutoffs were not optimized. Instead we chose cutoffs which are plausible (default cutoff of the geno2pheno[coreceptor] method and the proposed sensitivity limit of the phenotypic assay). Future work on larger datasets should focus on optimizing these cutoffs. The data set on which we based this study is too small for this purpose. However, when comparing the results to the predictions obtained with conventional bulk sequencing approaches it could clearly be shown that combining genotypic prediction with ultra-deep sequencing results in a fast and accurate alternative to phenotypic assays.

In addition to simple classification of the isolates into R5- and D/M-samples, we could also provide a quantitative output with this approach. This might be of minor interest when comparing against a phenotype as done in this work but is probably more important in the context of coreceptor antagonist administration. A patient with a very small minority of X4-viruses will perhaps respond differently to a coreceptor antagonist containing regiment than a patient with a majority of X4-viruses.

The main advantage of the massively parallel sequencing technology implemented in the

GS FLX-system is that it can reliably detect minor variants and generate clonal sequences. In the context of coreceptor usage, especially the first property seems to be of special interest because minor variants have been shown to be much more important in tropism determination than in HIV resistance testing.

There is room for improvements. First, as already discussed above one has to decide if the goal is to recreate the results of the Trofile assay or rather to reliably predict therapy outcome. In our opinion the latter is more important. Therefore the 5%-minority cutoff chosen in this project is probably not the best solution and should be adjusted when more treatment data become available. Furthermore, one might think about changing the dogma to exclude all patients from a therapy with coreceptor antagonists who harbor a dual-tropic or X4-virus. We found at least one sequence predicted as X4 in every isolate which means that if we do not use a minority cutoff none of our patients would be administered a coreceptor antagonist. This does not make therapeutic sense. Instead of trying to further increase the sensitivity of detecting minority populations we suggest to search for clinically relevant cutoffs. These cutoffs might not only be based on the number of minor X4-viruses but also depend on the fitness of these variants.

Another point to be addressed in the future is that we used a standard coreceptor prediction tool which was originally developed for bulk-sequenced isolates and not for ultra-deep data. A tool trained specifically on the latter kind of sequence data will most probably show enhanced prediction outcomes.

Finally, coreceptor antagonists are administered together with other antiretroviral compounds. We do not know if minorities of X4-viruses accumulate the same resistance mutations to these other drugs as the predominant R5-viruses. There might be the possibility that these variants have quite different polymerase-genes and therefore are resistant against a coreceptor antagonist but not against the other drugs in the backbone. Hence, a future goal is to sequence the whole virus with the ultra-deep sequencing technology in order to address this problem.

## 4.4 geno2pheno[454] - a Webservice to Infer Viral Tropism from 454-reads

In the previous section (4.3) we have shown how HIV-1 coreceptor usage can be inferred from viral genotype with massively parallel sequencing. Based on these results we have built up a website to deal with this task, the GENO2PHENO[454]-website. The webservice has been established at the end of June 2010. Until end of December 2010, it has been used more than 500 times which is remarkable considering the small number of labs using the 454-technology to determine HIV-1 coreceptor usage, so far.

As already mentioned several times in this chapter, the main problem of dealing with ultra-deep sequencing data is the vast amount of data. It is obvious that we cannot offer a website to which 100-300Mb will be uploaded and then processed. Thus, we have split our system into two parts: a pre-processor which is run locally in the lab and the website itself. The pre-processor handles the computationally expensive tasks such as the detection of the V3 loop and its alignment and creates a file which can be submitted to the website. The website then generates the predictions and reports them in different ways.

In the following we want to describe these two parts in more detail.

### 4.4.1 The Pre-processor

**Working with the program** The pre-processor is a JAVA-program that can be downloaded from the GENO2PHENO[454]-website. It is implemented in *JAVA* to be independent of the operating systems used in the different labs.

The program comes with a graphical user interface (GUI) displayed in Figure 4.20. It allows for loading an input file in standard flowgram format (.sff)- or fasta-like format (.fna, .fas, or .fasta).


Figure 4.20: Graphical user interface of the JAVA program used to pre-process the 454-data.

Furthermore, one can provide a file in fasta format containing the multiplex identifiers (MIDs, barcodes) used in the sequencing run. When provided, the system will demultiplex the reads by searching for the corresponding MIDs in the forward and reverse strains.

The tag file containing the MIDs is very easy to understand: The headers will be used as sample names while the data will be sorted according to the corresponding sequences of these headers. An example is given in Figure 4.21.

Using this list, the pre-processor would sort the data into three groups: one where sequences start with $tcagTTTT$, one in which sequences starting with $tcagAAAA$ and last but not least one junk-file containing all sequences that do not start with any of these motifs. $tcag$ is the key sequence usually used as the quality control of the 454 machine (see

```
>SampleA
tcagTTTT
>SampleB
tcagAAAA
```

Figure 4.21: Example for a tag-file.

also Section 4.1). If the samples have been trimmed already and do not start with such a key sequence, one can just remove it.

Optionally, one can choose a name for the output file, if not a file called *default.g2p* is generated.

Finally, pre-processing is started by clicking on the *"Prepare 454-data for geno2pheno"* button. When finished, the program will report this with a message popping up. It can then be closed and the generated file submitted to the website.

**Workflow** The program uses most of the approaches described in 4.2.1, specifically trimming and collapsing of the reads, the introduced sequence-to-profile alignment procedure, and detection of sequencing errors.

The workflow can roughly be described as follows:

**Input:**

- a set of reads $r_i$
- a sequence profile of V3 loops $P$
- a database-update quality cutoff $Q$

**Initialize:**
$D < V3loop, alignment >= \{\}$
$L < V3loop, alignment >= \{\}$

**Process:**
For every read:

1. translate the sequence into all six reading frames (three forward and three reverse)
2. check a database $D$ if one of the already aligned V3 loops is found within one of the six reading frames
3. if found, add the read to the list of the found V3 loops, go to 1) and process the next read
4. search within each frame for typical V3 motifs (e.g. CTR, CIR, AHC, AYC, ...)
5. if start and end-motifs are found in the same frame and if the cytosines encapsulating the loop are in a distance of 30-40 amino acids to each other, then:
   a) align this frame against the V3 profile
   b) compute the percentile score $p$ from the alignment score

   c) if $p < Q$, add the V3 loop and its alignment to the database $D$, go to 1) and process the next read

6. align each frame $f$ against the V3 profile
7. compute the percentile scores $p_f$ from the alignment scores
8. if $p_f < Q$, add the found V3 loop in frame $f$ and the corresponding alignment to the database $D$
9. otherwise add the V3 loop and its alignment of the frame with highest alignment-score to the low-scoring database $L$

We make several remarks at this point:

- The *database-update quality cutoff* $Q$ is an important parameter that should be selected carefully. If chosen too low, it rejects most reads which leads to the fact that almost every read will be aligned in all six reading frames and the database will be only used rarely. On the other hand, if the cutoff is chosen too high, it can happen that imperfect V3 loops are added to the database. Suppose we have a read that was not sequenced properly and terminated shortly before the end of the V3 loop. If, except for the missing 1 or 2 amino acids, this read resembles a typical V3 loop, then it will still obtain a good alignment score and be added to the database. If the next read processed is the same but with a complete loop, we would not align this read again but use the alignment stored in the database. This, however, lacks the remaining amino acids. Of course, this should not happen and therefore we suggest not changing this quality cutoff from 95% or checking the results very carefully.
- Reads with low-score alignments (high percentiles) are stored in the second database $L$ to allow the user to analyze them later on, too. On the website, these can still be screened out by selection of an appropriate percentile cutoff.

**Technical aspects** As mentioned before, the pre-processor is implemented in *JAVA*. *JAVA* version 1.6.0_18-b07 or newer is required for running the program. Otherwise the following error will appear. We note this cryptic error here as it does not suggest a problem with the installed *JAVA* version.

```
Exception in thread "main" java.lang.UnsupportedClassVersionError: Bad version number
in .class file
        at java.lang.ClassLoader.defineClass1(Native Method)
        at java.lang.ClassLoader.defineClass(ClassLoader.java:675)
        at java.security.SecureClassLoader.defineClass(SecureClassLoader.java:124)
        at java.net.URLClassLoader.defineClass(URLClassLoader.java:260)
        at java.net.URLClassLoader.access$100(URLClassLoader.java:56)
        at java.net.URLClassLoader$1.run(URLClassLoader.java:195)
        at java.security.AccessController.doPrivileged(Native Method)
        at java.net.URLClassLoader.findClass(URLClassLoader.java:188)
        at java.lang.ClassLoader.loadClass(ClassLoader.java:316)
        at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:280)
        at java.lang.ClassLoader.loadClass(ClassLoader.java:251)
        at java.lang.ClassLoader.loadClassInternal(ClassLoader.java:374)
```

Another issue is the amount of memory required. We have tested the program with GS FLX data containing up to 200,000 reads with an average read-length of about 350 base pairs as well as with Titanium runs of up to 1.5mio reads with an average read-length of 450 bp. The former runs needed 1 to 2 GB of memory while the Titanium runs needed up to 4 GB. If the provided memory is too low, the system will need to swap and run for days and weeks. Otherwise, a typical run is completed in 1 to 4 hours, dependent on the similarity of the individual reads. E.g. if all reads derive from the same patient, processing is usually very fast and can take as few as 15 minutes. On the other hand, if a lot of MIDs were used to sequence samples from many patients, the running time can go up to several hours.

### 4.4.2 The Website

**Input page** The input page of the webservice offers the user the different programs and files for pre-processing to download. At time of this writing, we offer the pre-processor, two tag-files, one containing the standard Titanium multiplex identifiers and one with a default tag, and a batch script (for Microsoft Windows users) that starts the pre-processor without the command line.

On the input page, the user can provide an identifier for the isolate, and submit the file generated by the pre-processor. In addition, a percentile cutoff can be selected determining which reads should be excluded from the final analysis.

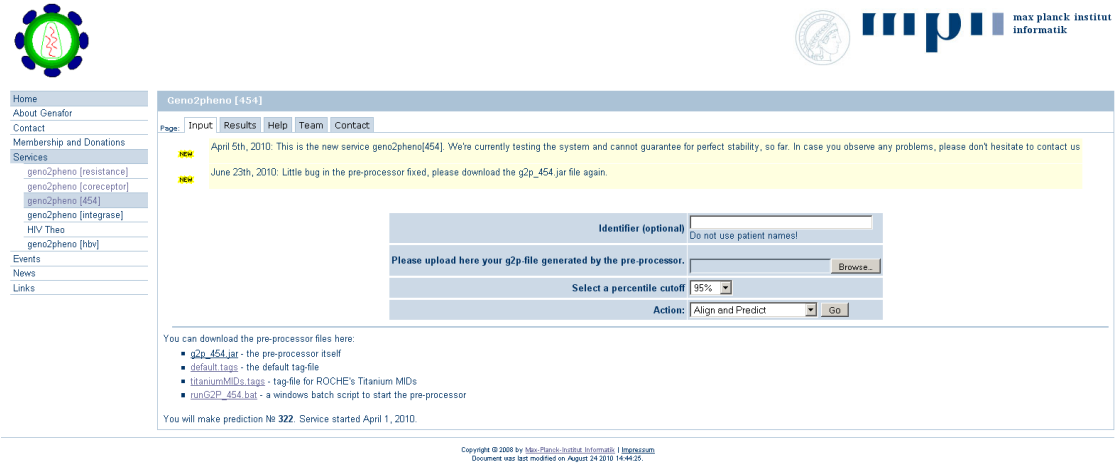


Figure 4.22: The input page of the GENO2PHENO[454]-website.

**Output page** The output page displays the results of the predictions in a compressed form. In the first part, general properties about the processed run are given:

- the name of the uploaded file
- the selected quality level
- the version number of the pre-processor
- how many reads were sequenced in total

- how many different V3 loops were found (based on amino acid sequences and including low-quality variants)
- how many different barcodes/multiplex identifiers were used

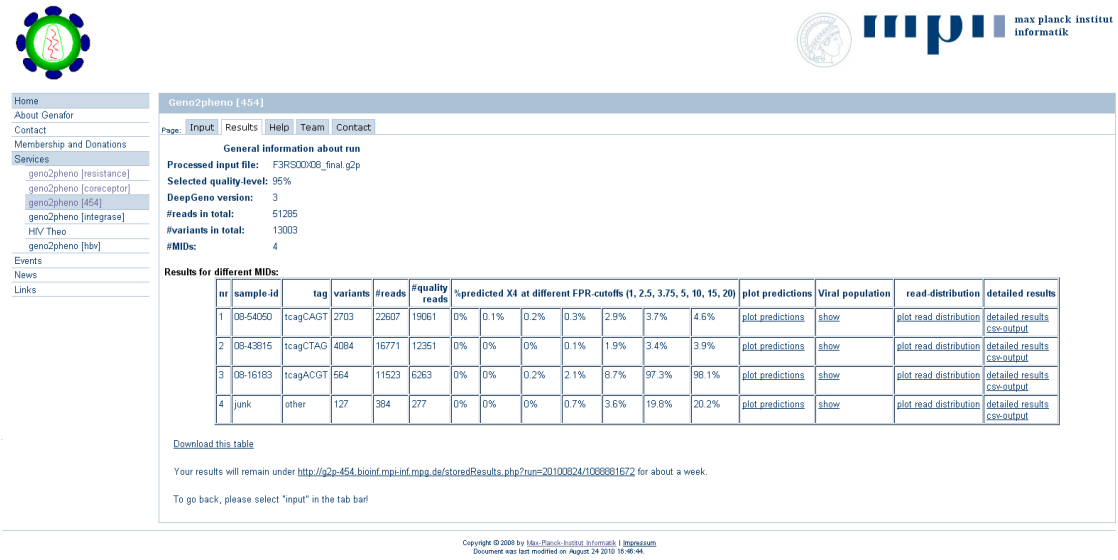


Figure 4.23: Result page of the GENO2PHENO[454]-website.

Below is a table with the results for the different MIDs (see Figure 4.23). Among others, it shows how many reads passed the quality-control and how many of the quality reads were predicted to be X4 for different FPRs. The table can be downloaded in csv-format.

The right-hand columns of the table contain links to further plots and more detailed results. Among them are the representations shown in Figures 4.11, 4.12, and 4.13. In addition, the distribution of the variants is shown as depicted in Figure 4.24. This plot describes how often the 20 most frequent variants occurred in the viral population, i.e. the leftmost bar denotes the prevalence of the most common V3 loop in this sample (in this case the most prevalent loop occurred in 33.7% of the reads).

Last but not least, a table containing more detailed results is provided (Fig. 4.25). It is available as an XML-output and a downloadable csv-file. The table contains the different variants, how often they occurred, the corresponding percentile values, and the predicted FPRs. In the example given in Figure 4.25, a V3 loop predicted to be X4 with a FPR of 10.9% was found in 5428 of the reads of the 454-run. The next row displays another sequence that was found 4077 times, however which would usually be screened out due to the high percentile-value of 0.988.

All results are stored under a specific URL for about a week so that one does not have to process the data over again. Instead the user is provided a link to a stored result page which can then be used to retrieve all the results again.

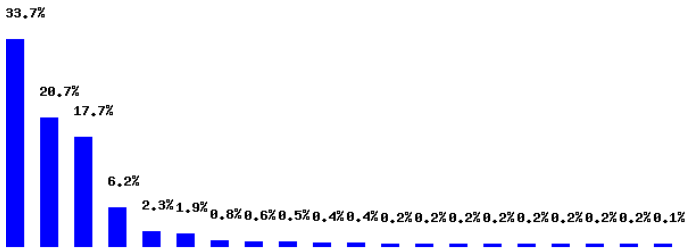


Figure 4.24: The prevalence of different V3 loop sequences in the sample. Each bar indicates one variant, the percentages above show how frequent these appeared in the viral population.

| READ-NR | OCCURRENCES | V3-LOOP | PERCENTILE | FPR |
|---|---|---|---|---|
| 379 | 5428 | CTRPSNNTRTSIRIGPGQVFYRTGDITGDIRKAHC | 0.626 | 10.9011627907 |
| 113 | 4077 | SSRRSSSTPDLPSQRLTGEHC | 0.988 | 4.21511627907 |
| 421 | 150 | AAKPDGGKRSVSVKIRCC*L*NNVC | 0.988 | 2.76162790698 |
| 204 | 126 | AAKPDGGKRSVSVKIRC | 0.988 | 2.03488372093 |
| 187 | 69 | CQPAQSQLGPLLRPKVAHSFFC | 0.988 | 7.55813953488 |
| 413 | 46 | CSPVSL*LGRSGWSCFFC | 0.988 | 0.726744186047 |
| 221 | 41 | CTRPSNNTRTSIRIGPGRVFYRTGDITGDIRKAHC | 0.607 | 3.92441860465 |
| 416 | 39 | CTRPPNNTRTSIRIGPGQVFYRTGDITGDIRKAHC | 0.679 | 6.83139534884 |
| 124 | 35 | CTRPSNNTRTSIRIGPGQVSYRTGDITGDIRKAHC | 0.824 | 8.13953488372 |
| 356 | 26 | CTRPSNNTRTSIRIGPGQVFYRTGGITGDIRKAHC | 0.712 | 6.97674418605 |
| 228 | 24 | CTRPSNNTRTSIRIGPGQVLYRTGDITGDIRKAHC | 0.74 | 23.6918604651 |
| 188 | 23 | CTRPSNNTRTSIRIGSGQVFYRTGDITGDIRKAHC | 0.784 | 4.06976744186 |
| 22 | 21 | CTRPSNSTRTSIRIGPGQVFYRTGDITGDIRKAHC | 0.702 | 5.08720930233 |
| 558 | 20 | CTRPSNNTRTSIRIGPGQVFYRTGDIAGDIRKAHC | 0.635 | 11.6279069767 |

Figure 4.25: Detailed prediction results.

# 5 Daily Routine - Working with the Bulk-sequenced V3 Loop

Within this chapter we will focus on the V3 loop and sequences derived with conventional population-sequencing methods. While other kinds of data discussed in Chapters 3 and 4 might be highly interesting and helpful for prediction of HIV-1 coreceptor usage, one should have in mind that such data is usually not generated in daily diagnostics. Because improving the daily routine is our major goal we have to find ways to improve predictions on bulk-sequenced V3 loop data, too.

In the first section (Section 5.1) we will show how the two most widely used prediction systems can be combined to a new prediction system. We will then focus on a topic which came up in the course of this work. A new phenotypic assays has become available and replaced the original Trofile assay: the Enhanced Sensitivity Trofile Assay (ESTA). The impact of this new assay with a higher sensitivity of detecting X4-viruses on prediction outcomes will be discussed in Section 5.2. The third section of this chapter will deal with different HIV subtypes which are thought to have different coreceptor usage prevalences. Deriving the HIV subtype information from the V3 loop is therefore appealing (Section 5.3). Subsequently, we will analyze in the fourth section of this chapter to which extent clonal data differs from clinically derived samples by comparing V3 loop sequences originating from different datasets (Section 5.4). In case of pronounced differences, we might have to think about choosing other datasets for training of the prediction system. Last but not least, we will round up this chapter with analyzing the HIV-1 coreceptor usage predictions with respect to therapy-outcome rather than the phenotypic results (Section 5.5).

## 5.1 From Two Make One - Combining the Two Most Often Used Systems

geno2pheno[coreceptor] is one of the two most widely used systems for prediction of HIV-1 coreceptor usage. Another very popular system, the WebPSSM-webserver developed by Mark Jensen at the Mullin's Lab in Seattle, Washington, is based on position specific scoring matrices (PSSMs).

These two systems have been compared in several publications (e.g. Low et al., 2007; Garrido et al., 2008b). In most cases, their overall prediction performances were quite similar but the individual predictions often differed. Because of this observation, several labs decided to use both systems in order to decrease the probability of falsely predicting an X4-isolate as R5. In this worst-case approach a sample is assigned to be X4, if any system predicts the respective sequence to be X4. Applying this approach, administration of maraviroc will only be considered, if both systems agree that it is an R5-virus. Of course, while the sensitivity to predict X4-viruses is increased, specificity is decreased. As

a consequence, more patients which could benefit from a regimen containing a coreceptor antagonist will not receive the drug.

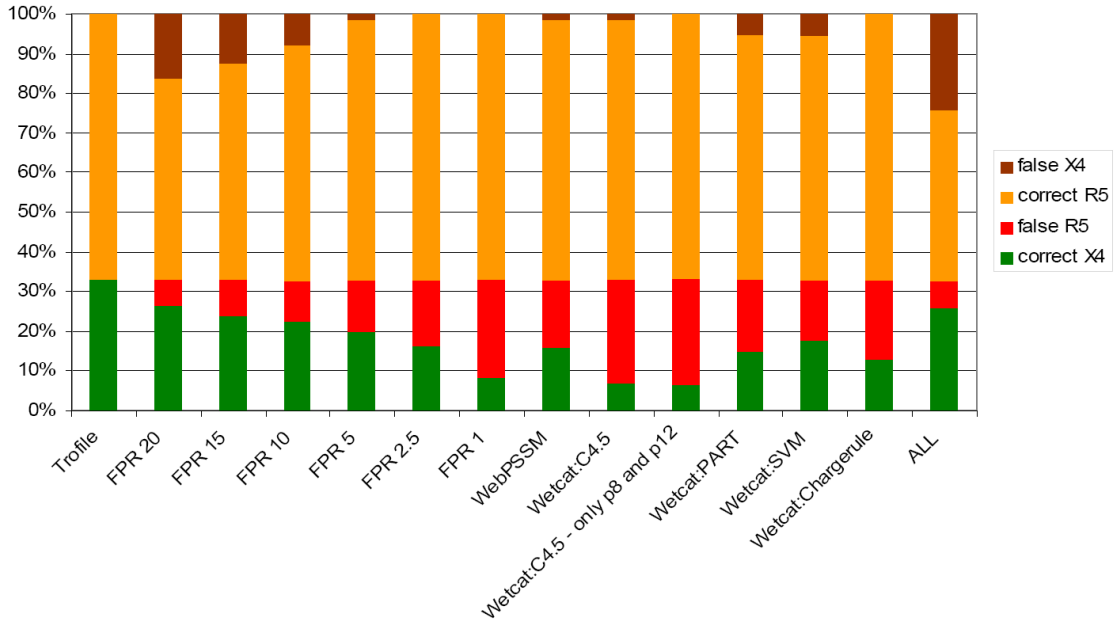


Figure 5.1: Agreement between different genotypic coreceptor prediction methods and Trofile (adapted from Obermeier et al., 2008).

A natural extension of this approach is to take further prediction systems into account. Obermeier et al. have combined the prediction results of seven methods (Obermeier et al., 2008). However, while the sensitivity of this approach rose to about 80%, the specificity decreased to just 65% (see Figure 5.1). The performance of geno2pheno[coreceptor] with a false-positive rate of 20% achieved comparable results (77% sensitivity, 70% specificity).

In another work by Chueca et al. (Chueca et al., 2009), the authors tried to relax the constraints a bit. Instead of calling a virus to be X4 if only one of four different methods predicts it to be X4, they took the majority call, i.e. if at least three systems agreed on R5 or X4 they used the respective prediction. The drawback of this is that they did not give a result in case that two systems voted for R5 and two for X4. While their results look impressive at first sight, they are not that convincing upon closer inspection: What should be done in daily diagnostics if no result is provided? Treat it as an R5- or an X4-virus? One probably wants to be on the safe side and thus would treat it as an X4-virus. As a consequence, the specificity of the method would decrease, too. The other option is to make a phenotypic test in these cases but this might lead to many samples being tested both with a genotypic and a phenotypic test. In contrast to a fast and inexpensive method, such an approach will increase both the costs and the turnaround time.

A common property of the described combination of methods is to use just the binary classification (R5 or X4) of the individual predictors but not to take their individual prediction scores into account. It does not make a difference for the combination methods if the classification of the individual predictors is assigned with high or low probability. In this section, we want to address this issue and discuss how prediction scores generated by

the two most often used prediction systems geno2pheno[coreceptor] and WebPSSM can be combined into a single score in order to improve prediction outcomes.

The work described in this section was carried out in collaboration with the British Columbia Center for Excellence in HIV (Vancouver, British Columbia, Canada), Pfizer Inc. (New York, USA), and Pfizer Research & Development (Sandwich, UK). Parts of it have been presented at the XVII International HIV Drug Resistance Workshop: Basic Principles and Clinical Implications 2008, Sitges, Spain (Moores et al., 2008).

### 5.1.1 Material & Methods

**Study samples** Samples from a subset of the HOMER cohort were used for training. These samples had a Trofile-phenotype (34 R5, 29 X4) and triplicate independent V3 genotype determinations were performed. Sequencing was achieved by independently amplifying each sample in triplicate. Two subsets of the MOTIVATE screening samples were used for evaluation. Both datasets were sequenced in triplicates as described before. The first dataset comprised 81 samples with 52 R5- and 29 D/M-samples whereas the second contained another 278 samples for which no tropism information was provided. Instead, predictions were evaluated in a blinded validation by Pfizer.

**Sequence handling** Coreceptor usage was predicted from genotype with the prediction tools geno2pheno[coreceptor] and WebPSSM. WebPSSM scores were provided by our cooperation partners at the British Columbia Center for Excellence in HIV (Vancouver, British Columbia, Canada). Because this method cannot handle sequence ambiguities, the genotypes were expanded into all possible combinations. The combination with the highest PSSM score (the most X4-like) was chosen. The expansion was not necessary for geno2pheno[coreceptor] which automatically computes the average-score of all ambiguous positions. For both prediction methods, the highest prediction score of the three replicates was finally chosen for further analysis.

**Generation of a combined score** In preliminary experiments we applied standard machine learning approaches like logistics regression and support vector machines for this task. However, they failed to provide good results. The reason for this remained unclear, it might have been the low number of training samples. A special problem here was that the predictions should have high specificities rather than high accuracies. Because we did not obtain useful results we switched to a more explorative strategy which was based on observations from visual inspection of the prediction scores.

Figure 5.2 depicts the scores from WebPSSM and geno2pheno[coreceptor] for the dataset used for training in this section. Green dots represent samples determined to be R5 by the Monogram Trofile Assay while red dots are D/M-isolates. The solid orange line represents the WebPSSM cutoff, while the dotted purple line shows the 5% false-positive rate cutoff of geno2pheno[coreceptor]. We chose the 5% cutoff rather than the default geno2pheno[coreceptor] cutoff of 10% here because taking the highest score from three replicates naturally increases the predicted FPR in comparison to sequencing and prediction just once. With a higher FPR we would consequently also increase the sensitivity but decrease the specificity, too.

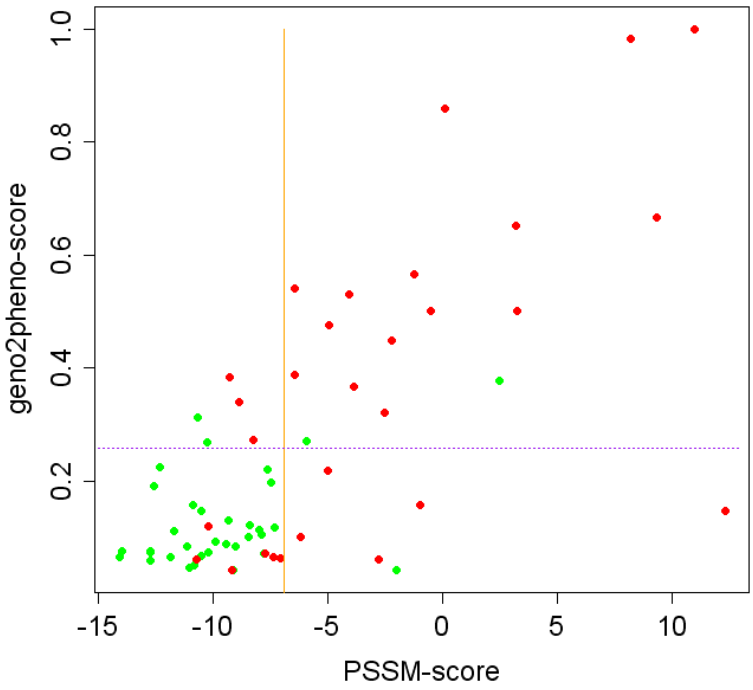


Figure 5.2: Prediction scores of WebPSSM and geno2pheno[coreceptor] plotted against each other for samples of the training dataset. The solid orange line describes the original WebPSSM cutoff to discriminate between R5- and X4-viruses whereas the dotted purple line depicts the 5% FPR cutoff of geno2pheno[coreceptor].

It can be seen in Figure 5.2 that there is a high correlation between the prediction scores which implies that for most samples the two systems agree. However, for some isolates there is disagreement between the systems: there are samples predicted to be X4 by geno2pheno[coreceptor] but R5 by WebPSSM (the ones above the purple line and left to the orange line) while others are predicted to be X4 by WebPSSM but not by geno2pheno[coreceptor] (right to the orange line but below the geno2pheno-cutoff). Applying the approaches described before (Chueca et al., 2009; Obermeier et al., 2008), one would predict all samples lying above the purple line or right to the orange line as X4 whereas samples occurring in the lower left corner are predicted as R5. In contrast, our idea is to generate a new single decision boundary which classifies samples on one side to one coreceptor and samples lying on the other side to the other coreceptor. The specificity of such a linear combination of the two prediction scores should have a specificity of at least 92%, the specificity of the 11/25-rule on this dataset. We tried to cover the space of possible decision lines comprehensively by computing linear separations through two points $P_1 = (pssm_1, g2p_1)$, $P_2 = (pssm_2, g2p_2)$ with the following requirements:

1. $P_1 \neq P_2$
2. $P_1$ and $P_2$ are located on the rectangle spanned by the points (min(pssm),min(g2p))

and (max(pssm),max(g2p)), where min(pssm) and min(g2p) represent the minimal WebPSSM and geno2pheno[coreceptor] scores occurring in the training dataset (max-scores analogous).

A line through these points can be written as:

$$y = m \cdot pssm + b \tag{5.1}$$

$$m = \frac{g2p_2 - g2p_1}{pssm_2 - pssm_1} \tag{5.2}$$

$$b = g2p_1 - \frac{g2p_2 - g2p_1}{pssm_2 - pssm_1} \cdot pssm_1 \tag{5.3}$$

Hence, a sample with WebPSSM score $pssm$ and geno2pheno[coreceptor] score $g2p$ would yield a prediction call:

$$P(pssm, g2p) = \begin{cases} \text{X4} & pssm \cdot m + b \leq g2p, \\ \text{R5} & \text{otherwise} \end{cases} \tag{5.4}$$

Each of the edges of the aforementioned rectangle was partitionated into 100 equispaced points so that we could test about 16,000 combinations. These were assessed for their ability to discriminate R5- from X4-samples. Combinations with a specificity of less than 92% were discarded. Among the remaining ones, the ones with the highest sensitivity were further selected.

### 5.1.2 Results

**Optimal linear combination** Several of the tested linear combinations separated the data equally well with a sensitivity of 72.4% and 94.1% specificity. Thus, we computed an "average" line by averaging the respective components of their points $P_1$ and $P_2$. The gradient and intercept of the resulting optimal decision line were:

$$\begin{aligned} m &= -0.0264585682974107 \\ b &= 0.0477976246568680 \end{aligned} \tag{5.5}$$

The resulting linear separation is depicted in Figure 5.3. Sensitivity and specificity were still at 72.4% and 94.1%. When testing the original methods we found sensitivities and specificities of 68.9% and 91.1% (WebPSSM) and 62.0% and 88.2% (geno2pheno[coreceptor], 5% FPR cutoff), respectively. We also analyzed two previously mentioned approaches to make use of the two prediction servers namely to take the worst-case scenario ($ANY$, a sequence is regarded to be X4 if any method predicts a sequence to be X4) or to use only samples in which two prediction systems agree ($AGREE$). In the first case, high sensitivity was achieved with 79.3%. However, due to the nature of the approach the probability of false positives was increased, too. Thus, specificitiy decreased to 85.3%. For the second approach, the two webservers agreed in 52 of the 63 samples (82.5%). Sensitivity for these 52 samples was 71.4% and specificity 93.5%. 8 of the remaining 11 samples had an X4 phenotype and 3 of them were determined to be R5.

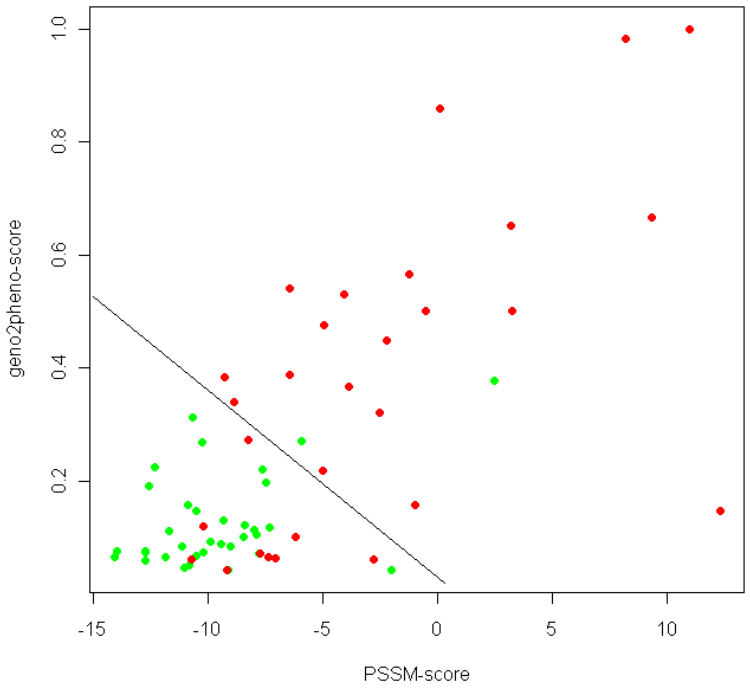


Figure 5.3: Prediction scores of WebPSSM and geno2pheno[coreceptor] plotted against each other for samples of the training dataset together with the determined linear separating plane. Points above the line are predicted to be X4 while points below the line are R5-predictions.

**Unblinded validation** For a first evaluation of our combined predictor, we had 81 samples from the MOTIVATE screening population with genotype and phenotype information (see Table 5.1). The two original methods WebPSSM and geno2pheno[coreceptor] both achieved an accuracy of 82.7%. While the sensitivity of geno2pheno[coreceptor] was 67.8%, WebPSSM reached a higher sensitivity of 75.0%. However, the specificity of geno2pheno[coreceptor] was a bit higher with 92.3% in comparison to the specificity of WebPSSM at 88.4%. The accuracy of the combined predictor was 88.9%. While having the same specificity as WebPSSM, the sensitivity increased to 85.7%.

We also compared against the two approaches *ANY* and *AGREE*. *ANY* achieved an accuracy of 81.5% with a sensitivity of 82.1% almost as good as the one of the combined predictor but with a much lower specificity of 82.7%. A prediction from *AGREE* was available in only 67 of the 81 samples (82.7%) with an accuracy for this subset of 89.5%. Sensitivity of *AGREE* to detect X4-viruses was just 73.9% but the specificity was with 97.7% very high.

**Blinded validation** Based on our first successful results, we wanted to validate our approach in a second blinded analysis. For this we obtained another set of sequences from the MOTIVATE study cohort but in order to be totally unbiased we received no phenotypic data. Instead, we used our combined method to predict the tropism of these isolates and

| *method* | *TP* | *FN* | *FP* | *TN* |
|---|---|---|---|---|
| WebPSSM | 21 | 8 | 6 | 46 |
| geno2pheno[coreceptor] | 19 | 10 | 4 | 48 |
| combined method | 24 | 5 | 4 | 48 |
| ANY | 23 | 6 | 9 | 43 |
| AGREE | 17 | 6 | 1 | 43 |

Table 5.1: Performance of the different methods on data from the MOTIVATE cohort. TP: true positives, correctly predicted X4-viruses, FN: false negatives, X4-viruses predicted as R5, FP: false positives, sequences falsely predicted to be X4, TN: true negatives, correctly predicted R5-viruses

sent the predictions to Pfizer where an independent evaluation was performed. The final results show the performances of the WebPSSM method, geno2pheno[coreceptor], and the combination of both (see Table 5.2).

| | *BLINDED SET* | | | | *UNBLINDED SET* | | | |
|---|---|---|---|---|---|---|---|---|
| *method* | *TP* | *FN* | *FP* | *TN* | *TP* | *FN* | *FP* | *TN* |
| WebPSSM | 87 | 29 | 28 | 134 | 552 | 151 | 293 | 920 |
| geno2pheno[coreceptor] | 71 | 45 | 12 | 150 | 441 | 262 | 95 | 1118 |
| combined method | 84 | 32 | 19 | 143 | 518 | 185 | 161 | 1052 |

Table 5.2: Performance of the different methods in the blinded analysis. TP: true positives, correctly predicted X4-viruses, FN: false negatives, X4-viruses predicted as R5, FP: false positives, sequences falsely predicted to be X4, TN: true negatives, correctly predicted R5-viruses

geno2pheno[coreceptor] achieved a sensitivity of 61.2% and a specificity of 92.6% whereas the WebPSSM method reached a sensitivity of 75% but only a specificity of 82.7%. The combined method almost reached the same sensitivity as the WebPSSM-method with 72.4% but showed a much higher specificity of 88.3%. Since the combined method almost reaches the sensitivity of WebPSSM and the specificity of geno2pheno[coreceptor], it seems that it combines the advantages of both methods thereby being more robust.

Recently, also the complete MOTIVATE screening dataset was provided to us containing 1916 genotype-phenotype pairs with 703 X4- and 1213 R5-samples. When retesting our method on these samples, WebPSSM achieved again the highest sensitivity (75.5%) but also the lowest specificity (75.8%). geno2pheno[coreceptor] in contrast showed the highest specificity with 92.1% but lowest sensitivity of the three tested methods (62.7%). Last but not least, the combination of the two systems generated predictions with intermediate performance. With 73.7%, sensitivity was almost as high as for WebPSSM whereas its specificity was much higher (86.7%) but not as high as for geno2pheno[coreceptor]. Of note, we did not have any numbers for the approaches *ANY* and *AGREE* in the blinded validation, but we checked them for the full dataset. *AGREE* yielded a very good performance with 75.9% sensitivity and 92.6% specificity. However, unfortunately the method could not provide any result for 20.3% of the samples. *ANY* on the other hand showed a sensitivity of 80.8% with a specificity of 73.8%. The

scores of geno2pheno[coreceptor] and WebPSSM together with the corresponding decision lines (orange: WebPSSM, purple: geno2pheno[coreceptor] with a FPR cut-off of 5%, black: combined predictor) are shown in Figure 5.4. Red points mark samples determined to be X4 by Trofile while the R5-viruses are colored in green. All points above the geno2pheno[coreceptor] decision line or right to the one of WebPSSM are predicted to be X4 by *ANY* while the approach *AGREE* can be found by looking at the points below the geno2pheno[coreceptor] *and* left to the WebPSSM line (R5-predictions) as well as above the one of geno2pheno[coreceptor] *and* right to the WebPSSM line (X4-predictions). All other points are not scored by this approach.

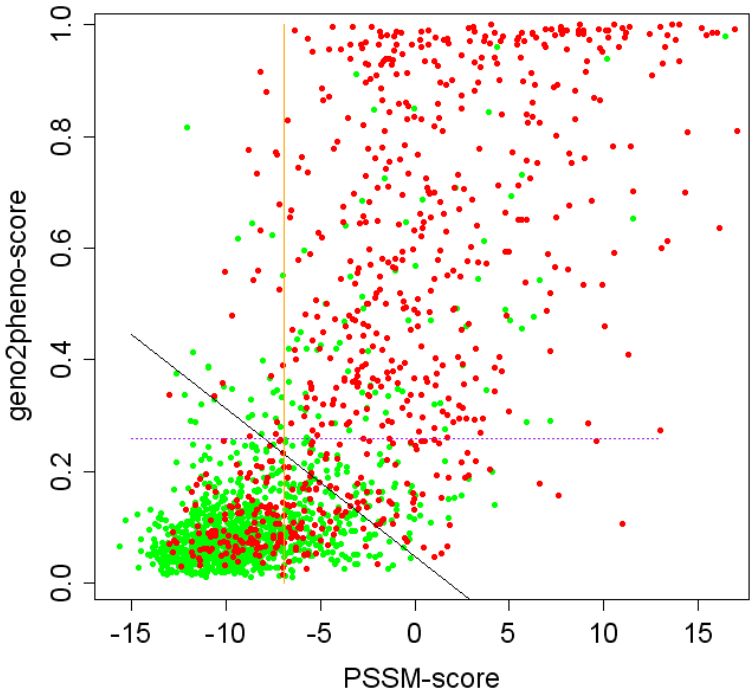


Figure 5.4: Prediction scores of WebPSSM and geno2pheno[coreceptor] from samples of the MOTIVATE screening population. The linear combination of the two systems is marked by the solid black line. Points above the line are predicted to be X4 while points below the line are R5-predictions. The solid orange line describes the original WebPSSM cutoff and the dotted purple line depicts the 5% FPR cutoff of geno2pheno[coreceptor].

### 5.1.3 Conclusions

In this analysis we could show how the scores of the two most prominent web systems for coreceptor usage prediction can be combined. Our results suggest that this combination indeed can improve the prediction quality. Despite training on a small dataset, the obtained decision boundary showed very good results on the MOTIVATE dataset. The validity of the described approach has been underlined with the blinded validation.

In general, the original methods used for training, geno2pheno[coreceptor] and WebPSSM, yielded similar results displaying sensitivities in the range of 60% to 70% and specificities of around 90%. In this analysis, however, they differed substantially. While the results of geno2pheno[coreceptor] were still in the same range, WebPSSM showed higher sensitivities with much lower specificities. This is probably due to the fact that triplicate sequencing was used and the most X4-like prediction was chosen. Thus, a more conservative cutoff should be selected for the prediction methods. In case of predictions from geno2pheno[coreceptor] we corrected for this by lowering the FPR cutoff to 5%. With respect to the WebPSSM method, our cooperation partners restrained from doing this in the blinded analysis. For the complete dataset we therefore also checked the results with a higher PSSM cutoff (data not shown). Surprisingly, with a cutoff reflecting the specificity of geno2pheno[coreceptor] the sensitivity of WebPSSM was reduced by more than 7 percentage points (sensitivity of about 55%). For the generation of the combined predictor this had no effect because the raw scores were used so that the method was independent of a specific cutoff.

The results of the combined predictor were lying in between the ones of its founder methods. Sensitivity was almost as high as the one of WebPSSM while specificity remained high with values similar to the one of geno2pheno[coreceptor]. When changing the cutoffs of the two web systems so that they reflected the specificity of the combined predictor, the latter still outcompeted them by at least 3 percentage points in sensitivity.

The approaches *ANY* and *AGREE* also showed impressive results when focusing on the exclusion of patients with an X4-virus. However, for daily routine they simply excluded too many patients, i.e. their specificity was quite low, in general. In case of *AGREE* one has to note the fact that every sample that cannot be predicted by the method has to be considered as a potential X4-isolate. Thus, about 20% of the patients tested would be deprived of coreceptor antagonists by default. Furthermore, the question of where to set the cutoffs is also a problem in these approaches. While there is already a major debate in the community on where to set the FPR cutoff of geno2pheno[coreceptor] these approaches would also need to define a second optimal cutoff for the PSSM score. To find the perfect combination of cutoffs that are also accepted in the community is certainly not an easy problem.

Nevertheless, the presented approach has some drawbacks: First of all, as we already mentioned, the sequence data we used here came from triplicate sequencing. While this strategy has been shown to improve prediction quality in some works, it is currently not the method of choice in daily diagnostics. We have tested the computed decision line on data that was generated with single sequencing but the performance did not improve over the geno2pheno[coreceptor] system. This might be because the maximum score of the three replicates was used which is higher than arising from single sequencing. As a consequence, training should be repeated based on a dataset where samples were sequenced just once.

A more practical problem is that we are dealing with two independent web-systems. While the described method yields interesting results a useful implementation would probably need to be offline. This is because the WebPSSM method cannot deal with sequence ambiguities. Therefore, our cooperation partners had to expand the sequences into all possible combinations before submitting these sequences to a local version of the WebPSSM

method. E.g. if there were two ambiguous positions with two possible residues respectively, they tested 4 different sequences. From a methodological point, this is not very problematic. However, the number of clones raises exponentially with the number of ambiguous positions. This makes the predictions very slow and not easy to send over the internet.

In this context, another major problem might arise in the future as more and more data will be generated with next generation sequencing technologies. While it would definitively be interesting to see the effect of the combined method on this type of data, the current way of combining the scores would require a client-based implementation of the two web systems as well as of their combination. While we have developed routines for handling next generation sequencing data with the geno2pheno[coreceptor] software, there are no alignment routines for the WebPSSM method which would have to be different from the one implemented in geno2pheno[coreceptor].

Last but not least, one has to keep in mind that with every update of one of the two prediction systems, one would also need to train a new combined method. This would add another level of complication in terms of implementation and evaluation.

## 5.2 Better Phenotype - Worse Genotype?

For most studies in the realm of HIV-1 coreceptor usage as well as for most analyses shown in this thesis, the Monogram Trofile Tropism Assay is used as reference. Due to the fact that the assay has been used in all clinical trials of the Pfizer drug *maraviroc* and also in the respective studies of *vicriviroc* (Schering Plough), the assay is widely considered as being the gold standard.

In general, Trofile is advertised to have been successfully run for more than 40,000 samples. However, successfully performing an assay does not necessarily mean that it is also always correct. In fact, there are several reasons to doubt correctness. For example, if an X4-virus is prevalent only in specific compartments of the body, even the perfect assay will not find the virus if it is based on a sample from a different compartment.

That Trofile is not always correct has also been discussed in a work by Skrabal et al. (Skrabal et al., 2007) in which two commercially available phenotypic assays were compared, the Monogram Trofile assay and the Tropism Recombinant Test by Viralliance. From a subset of 93 antiretroviral samples of the HOMER cohort, a phenotypic result was available from both assays in 74 cases. The overall concordance was only 85.1%. Using geno2pheno[coreceptor] similar results were achieved with 86.5% and 79.7% concordances with the Trofile and TRT assays, respectively. It is unlikely that any of the two assays tested was always correct because the discordances did not show significant trends that could be attributed to differences between the two phenotypic approaches or to patient characteristics (Skrabal et al., 2007).

In addition to this work, some studies of the maraviroc clinical trial also pointed to possible issues with the phenotype. For example, in the *MERIT* studies it was observed that in about 8% of the patients, tropism changed between screening and baseline, i.e. the measurement at the time of therapy start. These samples were considered to have minorities of X4-variants at the limit of detection of the assay. Unfortunately, while the MERIT-study failed to reach the study-goal, re-analyses showed that when these samples were excluded, the non-inferiority of maraviroc compared to efavirenz would have been demonstrated and the drug would have been approved also for therapy-naïve patients (Cooper et al., 2010). Thus, in the course of this work a new phenotypic test has replaced the original Monogram Trofile Tropism Assay: the *Enhanced Sensitivity Trofile Assay (ESTA)*. The enhanced sensitivity version of Trofile allows an average 30-fold improved detection of minor CXCR4-using variants in env clone mixtures. It is claimed to have 100% sensitivity at detecting X4-viruses with a prevalence of 0.3% of the viral quasispecies (Coakley et al., 2009).

The question arising from this development was: If the original phenotypic assay had already a better sensitivity in detecting X4-viruses in comparison to standard bulk-sequencing methods, how would the performances of genotypic coreceptor usage prediction change with respect to the new ESTA? A more sensitive phenotype should imply that the sensitivities of prediction methods may go down once more.

In the following brief analysis we will investigate if the number of X4-viruses increases when using the enhanced Trofile assay and compare the performances of coreceptor usage prediction with geno2pheno[coreceptor] in samples screened either with the old or with the new phenotypic assay.

The work described in this section was carried out in collaboration with the HIV-GRADE consortia. It has been presented at the 7th European HIV Drug Resistance Workshop 2009, Stockholm, Sweden (Thielen et al., 2009).

### 5.2.1 Material & Methods

**Dataset** 738 isolates screened for administration with the coreceptor antagonist maraviroc were used for analysis. All samples were phenotyped with the Monogram Trofile Tropism Assay, 619 (83.9%) with the original Trofile assay (*standard dataset*), and 119 (16.1%) with the enhanced sensitivity version (*enhanced dataset*). The V3 loop was determined for all samples with conventional bulk sequencing methods.

**Performance Assessment** Coreceptor usage was determined from viral genotype with the geno2pheno[coreceptor] tool. Prediction outcomes on the two types of data were compared using the false-positive rate cutoffs available on the geno2pheno[coreceptor] webpage (1%, 2.5%, 5%, 10%, 15%, and 20%). As measures, the sensitivity and specificity to detect X4-viruses and the area under the ROC curve were used. The prevalences of X4-viruses in the two datasets were compared using Fisher's exact test.

### 5.2.2 Results

**Phenotype distribution** 619 samples were phenotyped with the original Trofile assay. 414 (66.9%) of these were determined to be R5 and 205 (33.1%) had an X4- or D/M-phenotype. In the dataset screened with the enhanced Trofile assay, the proportion of D/M- and X4-samples increased to 39.5% (47 of 119) while 60.5% were phenotyped to be R5. Possibly due to the low sample size, the number of X4-viruses was not significantly higher ($p = 0.1085$) in the dataset screened with ESTA than in the dataset screened with the original Trofile assay.

**Prediction outcomes** Table 5.3 lists the sensitivities to predict X4-viruses on data phenotyped with the old and the new assay. It is apparent that the sensitivity is lower on samples phenotyped with the new assay than on samples screened with the original Trofile assay, especially at the lower false-positive rate cutoffs. For example, using geno2pheno[coreceptor], the sensitivity of predicting X4-viruses was 63.4% on the dataset phenotyped with the original Trofile and 59.6% for samples determined to be X4 by the enhanced sensitivity Trofile assay at the geno2pheno[coreceptor] false-positive rate setting of 10%.

| | *1%* | *2.5%* | *5%* | *10%* | *15%* | *20%* |
|---|---|---|---|---|---|---|
| **original Trofile** | 20.9% | 40.5% | 55.1% | 63.4% | 67.8% | 75.6% |
| **enhanced Trofile** | 6.3% | 27.7% | 40.4% | 59.6% | 68.1% | 72.3% |

Table 5.3: Sensitivity of geno2pheno[coreceptor] to detect CXCR4-using variants at different false-positive rate cutoffs in two datasets screened with different assays.

However, when comparing the specificities, to a lesser degree, one can observe the opposite effect: the specificities tend to be higher on data generated with ESTA (see Table

| | *1%* | *2.5%* | *5%* | *10%* | *15%* | *20%* |
|---|---|---|---|---|---|---|
| **original Trofile** | 99.8% | 99.3% | 96.6% | 89.6% | 82.4% | 77.1% |
| **enhanced Trofile** | 98.6% | 98.6% | 97.2% | 91.7% | 84.7% | 77.8% |

Table 5.4: Specificity of GENO2PHENO[CORECEPTOR] to detect CXCR4-using variants at different false-positive rate cutoffs in two datasets screened with different assays.

5.4). At the aforementioned FPR-setting of 10%, specificity increased from 89.6% with respect to the standard Trofile to 91.7% for isolates phenotyped with ESTA. In general, one can observe that sensitivities and specificities at one FPR cutoff with respect to the original Trofile were similar to the ones with the next higher false-positive rate cutoff on ESTA data.

The combination of lower sensitivities with slightly higher specificities made the comparison of the prediction outcomes a bit complicated. Therefore, we also computed the ROC curves of the predictions on the two datasets and compared them. The blue curve in Figure 5.5 displays the ROC curve generated from the samples phenotyped with the original Trofile assay while the red curve depicts the prediction performance on the ESTA-dataset. Overall, the results were fairly similar with a slightly better performance of GENO2PHENO[CORECEPTOR] on the original Trofile dataset at high specificities. The area under the ROC curve, on the other hand, was slightly higher for predictions on the ESTA-dataset (0.845 vs. 0.838).

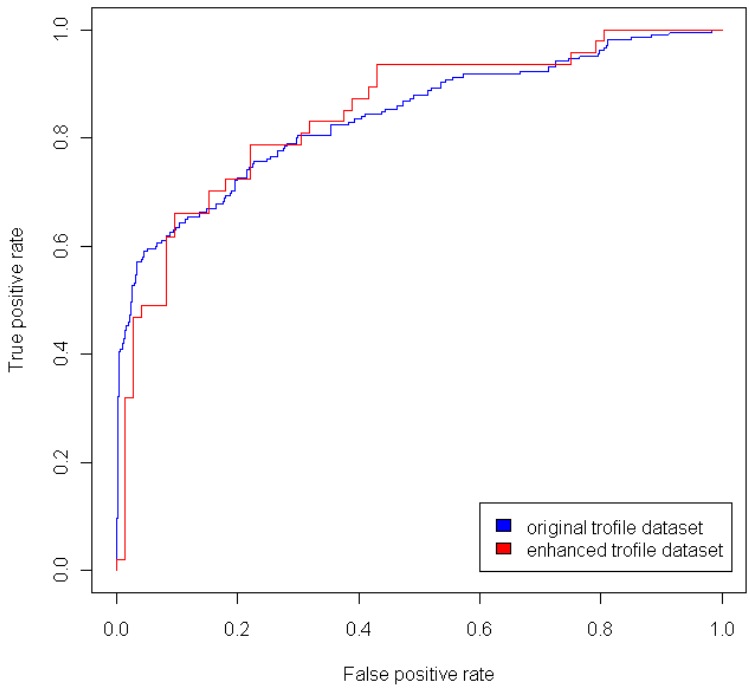


Figure 5.5: ROC curves of GENO2PHENO[CORECEPTOR] predictions on data phenotyped with the original (blue) or the enhanced sensitivity Trofile assay (red).

In order to obtain similar prediction probabilities with ESTA as with the old assay, one could adjust the FPR-levels such that the specificities are comparable to the ones used before. We analyzed which FPR-rates would be necessary to achieve a specificity on the ESTA-samples that is at least as high as with the FPR cutoffs described above. For the 1% and 2.5% FPR-levels, having a specificity of 99.8% and 99.3% on the samples phenotyped with the original Trofile we would need to use an FPR cutoff of 0.145%. With this the specificity on ESTA-samples would be at 100% but with a sensitivity of only 2.1%. For the 5% cutoff with 96.6% specificity on the dataset screened with the original Trofile, we would use an FPR of 6% instead corresponding to a sensitivity and specificity of 46.8% and 97.2%, respectively. The 10%-FPR cutoff would be replaced with a 12% FPR having a sensitivity of 65.9% and a specificity of 90.2% which would be both higher than the corresponding results on the original Trofile data. The 15% and 20% FPR cutoffs would be changed to 15.6% and 22.5% levels achieving sensitivities of 70.2% and 78.7% at specificities of 83.3% and 77.7%, respectively.

### 5.2.3 Discussion

Our results indicate that the new enhanced sensitivity Trofile assay detects more CXCR4-using viruses than the previous version. However, possibly due to the low sample size of ESTA-results, we could not observe a significantly higher number of X4-viruses. One of the reasons for this might be a slightly different patient collective. When maraviroc came to market, at first only highly treatment-experienced patients were screened for X4-viruses. These were phenotyped with the original Trofile assay. When the ESTA-assay was released, also less treatment-experienced patients at an earlier stage of infection were phenotyped for coreceptor antagonists. Because the more an infection is advanced the higher the probability of detecting an X4-virus, one would expect less X4-viruses in the samples screened with the enhanced Trofile assay.

However, the results on this cohort also demonstrate that the increased sensitivity of the phenotypic assay does not necessarily imply a decreased competitiveness of genotypic predictions. It is certainly true that the sensitivities decrease when the same FPR cutoffs are used. However, as our results clearly show, the corresponding specificities increased, too. Because the ROC curve of the original Trofile assay did not outcompete the one of the enhanced Trofile assay, we do not see any reason why geno2pheno[coreceptor] should perform worse than before. Instead, we could show that by adjusting the FPR of geno2pheno[coreceptor], prediction results became at least comparable to the ones of the original Trofile assay.

A question arising from these results is if the ESTA is really an improvement with respect to the original Trofile assay. Our results do not support this viewpoint. Instead, they suggest that by enhancing the sensitivity of the assay its specificity has been decreased. This might be due to a lower threshold used for the relative light units (RLU) in the assay. These measure the luciferase activity which is emitted upon successful infection of a cell in the assay. However, so far it has not been reported in what way the assay has been improved. Therefore it is hard to find further reasons for this problem and we can only imagine this as one possible explanation.

On the other hand, one also has to admit that the sizes of the datasets used here were

limited. Therefore, the results should be interpreted with care. Specifically, we would not suggest to rely exactly on the adjusted FPRs. Nevertheless, a trend that can be seen over all adjusted FPRs is that they are usually 1% to 3% higher than the corresponding ones used on traditional data. Thus, with an increase of 3% one should obtain good results in most cases. Further analyses with larger datasets should allow indentifying more precise cutoffs in the future.

## 5.3 Coreceptor Usage Prediction in Different Subtypes

Several reports have shown that the prevalence of R5- and X4-viruses is different among subtypes. Some subtypes such as A, C, or G have been demonstrated to present R5 tropism more frequently (Tscherning et al., 1998) than subtype B whereas e.g. subtype D samples are known to be more often of X4 tropism (Huang et al., 2007).

The reasons for these differences are still unclear. So far, not enough data could be analyzed to give more than hints. Another related question is to which extent different subtypes can affect the performances of prediction systems. Because the different genotypic prediction systems are trained mainly on subtype B data, some people believe that such systems would perform poorly when tested on non-B subtypes. As a consequence, the authors of the WebPSSM-method generated an additional scoring matrix specifically designed for subtype C samples (Jensen et al., 2006).

In our opinion this is not the right way to handle this problem due to several reasons:

1. We would need to train a separate model for each subtype and circulating recombinant form. The problem is that for most subtypes or recombinants we will probably never have enough training data so that we would either need to use prediction models of another subtype or to decline a prediction.

2. We think that the properties defining tropism are the same among different subtypes and that tropism is defined by the genotype. For example, the 11/25 rule is generally believed to work in all subtypes. We do not see any reason why this property should not be shared by other mutations defining coreceptor usage. Instead, we believe that certain subtypes evolve X4-features more often than others. This might be due to different wild-type residues that can be mutated more easily (e.g. because of a lower genetic barrier similar to the K65R NTRI resistance mutation occurring more often in subtype C samples, compare with Invernizzi et al., 2009) but also because of other properties not encoded in the envelope protein (e.g. higher mutation rates due to mutations in the reverse transcriptase).

3. It is complicated to deal with recombinants. Suppose we have a sample from the recombinant CRF02_AG. Do we need to have an AG-predictor for this? Or can we use a method trained on subtype A because the envelope region of this recombinant is almost entirely A? Furthermore, do we need different prediction models for subsubtypes A1 and A2?

4. Subtypes are most often defined by their *pol*-sequences. However, as can be seen for some recombinant forms the *pol*-subtype does not necessarily have to be the same as the one of the envelope protein. Thus, using the *pol*-subtype is of course not the best solution if you are looking at properties defined in the envelope of the virus.

Based on these considerations, we suggest to concentrate only on the V3 loop and thus we will define a *V3-subtype*. Instead of training a prediction model for each subtype we can then incorporate the V3-subtype directly into the prediction system as an additional feature.

In this section, however, we will concentrate on the prediction of the *V3-subtype* and on the question if prediction systems really perform worse on samples not being of subtype B as suggested by others (e.g. Jensen et al., 2006).

The work described in this section was carried out in collaboration with the HIV-GRADE consortium. It has been presented at the 7th European HIV Drug Resistance Workshop 2009, Stockholm, Sweden (Thielen et al., 2009).

### 5.3.1 Material & Methods

**Datasets** For training and testing of the subtype predictor we downloaded all V3-sequences with available subtype information from the Los Alamos HIV Sequence Database. To avoid epidemiological bias of similar sequences we restricted our dataset to one randomly chosen sequence per patient. In addition, due to the limited number of sequences for some subtypes, our analysis was just based on the following subtypes or recombinant forms: A, CRF01_AE (AE), CRF02_AG (AG), B, C, D, F, G, and O. The number of available sequences is displayed in Table 5.5.

| *subtype* | *number of sequences* |
|---|---|
| A | 3056 |
| AE | 1650 |
| AG | 933 |
| B | 9556 |
| C | 3752 |
| D | 842 |
| F | 310 |
| G | 484 |
| O | 137 |
| total | 20720 |

Table 5.5: Number of sequences used for training and evaluation of the subtype predictor.

Because the envelope part of the recombinant form AG is of subtype A, we pooled these two clades. Furthermore, subsubtypes A1 and A2 as well as F1 and F2 showed only marginal differences in initial experiments and were therefore merged to a subtype A and a subtype F dataset, respectively.

The second dataset used in this analysis consists of 738 clinically derived samples screened for maraviroc administration in different labs in Germany. These samples were sequenced with conventional population-sequencing and phenotyped with the original Monogram Trofile Tropism Assay. The dataset served as a further validation set for the subtype predictor and for evaluating the performance of coreceptor usage prediction with GENO2PHENO[CORECEPTOR] on different clades.

**Inferring the subtype from the V3-genotype** The V3 loop is a very short region of around 35 amino acids. This fact poses a problem when inferring viral subtype from it. Because HIV-1 can have different subtypes in different parts of its genome, one can of course not

hope to infer the correct subtype or recombinant form of the whole virus from such a short region.

Nevertheless, experience has revealed that, at least to some extent, subtype-information can be inferred from the V3 loop. We have seen that the frequencies of some residues at specific positions vary substantially between different subtypes. For example, position 13 of the loop has a histidine in the consensus of subtype B samples whereas the consensus of subtype C displays there an arginine and subtype D viruses mostly have a proline at this position. Unfortunately, while this is true for the consensus sequences, there are also many subtype B samples with e.g. a proline at position 13 or subtype D samples with an arginine there. While sequences of a subtype share some properties, the picture is not as clear cut as in other parts of the viral genome because the V3 loop is highly variable and very short. As a consequence we hypothesize that the subtype of the loop cannot be inferred just by comparing a sequence against the consensus or a reference as often done for the polymerase region.

Instead we constructed a subtype-predictor in two main steps: First, we wanted to find a measure of similarity to a specific subtype. This similarity score should not be based on a comparison to a single specific reference sequence but to the whole set of samples available for training. For this purpose, we decided to use *one-class SVMs* that are often used for outlier detection.

To generate these models we aligned and encoded our sequences as regularly done in GENO2PHENO[CORECEPTOR], i.e. the sequences were sequentially aligned against the reference profile of GENO2PHENO[CORECEPTOR] and then encoded with binary indicator encoding. All training samples of the same subtype are then used to train linear one-class SVM models using the *LIBSVM* implementation in the *R*-package *e1071* (Chang and Lin, 2001).

For the second step of the prediction, a sequence to be subtyped is submitted to each one-class SVM model. The resulting probability scores are then used to encode for a second set of support vector machines, in this case, two-class SVMs in which the prediction score of each one-class model is used as an input feature. Again, for every subtype a support vector machine model is generated in one vs. all fashion, i.e. all samples not of the particular subtype are treated as negative examples while the ones of this subtype are labeled as positives. Again, the *LIBSVM* implementation of the *R*-package *e1071* (Chang and Lin, 2001) was used for training and validation.

When determining the V3-subtype of a given sequence, the sequence is encoded and the one-class SVM prediction scores are computed. These are then input to the two-class SVM prediction models which compute a probability for every subtype. Finally, the subtype of the model providing the highest probability score is chosen as the predicted subtype. We will term this two-step approach as *TWO-STEP*.

**Evaluation of the subtype-predictions** The subtype-prediction models were evaluated with stratified ten-fold cross validation on the Los Alamos dataset. Cost-factors $c$ and $\nu$-values were optimized with a grid-search in initial experiments on a subset of 10% of the data. Optimal factors were found to be $\nu = 0.25$ and $c = 0.1$. Predictions were compared against two other approaches: a) using multi-class SVMs (*MULTI*) encoded in the same way as the one-class SVM models described before, and b) with BLAST searches (Altschul

et al., 1990) that predict subtype on the basis of the i) best hit (*BLAST-1*), ii) the subtype most often occurring in the ten (*BLAST-10*), and (iii) 100 best hits (*BLAST-100*). In case that two subtypes occurred equally often, one was chosen randomly.

All approaches were trained and evaluated in each cross-validation fold on the same set of samples. The multi-class SVMs were trained using cost-factors $c = 0.001$, $c = 0.01$, and $c = 0.1$. The BLAST searches were performed against a database generated by the training samples of the respective cross-validation fold.

Subtype prediction models trained on the whole dataset with *TWO-STEP* were further tested on the clinical dataset against the REGA-subtyping tool (de Oliveira et al., 2005). Due to the limited amount of data, we compared only subtype B versus non-B predictions.

In order to investigate if GENO2PHENO[CORECEPTOR] performs worse on non-B than on subtype B samples, the subtypes of the samples in the clinical dataset were predicted and two datasets encompassing B or non-B samples were built. Coreceptor usage prediction of GENO2PHENO[CORECEPTOR] was then analyzed for these two sets individually. ROC-analysis was used to compare the two sample sets. The R-package *ROCR* (Sing et al., 2004) was used for evaluation.

### 5.3.2 Results

**Performance in cross-validation analysis** We evaluated the subtype-predictor with ten-fold cross validation on the Los Alamos dataset. First, one-class SVM models were generated for each subtype. This was done using the subtype-sequences provided in the training step of each cross-validation fold individually.

The performances of these models to detect sequences of the respective subtype are shown in the red ROC curves in Figure 5.6. The blue curves display the corresponding *TWO-STEP* predictions after the second step of the subtype-prediction system. These clearly outperformed the models generated in the first step of the approach in all subtypes and generally showed very promising results.

All Support Vector Machine models of *TWO-STEP* estimate the probability whether a sequence belongs to a specific subtype but they do not compute which subtype is the most probable. Thus, we chose to use the model displaying the highest probability score for subtype prediction. Using this approach we could correctly classify 86.66% of the 20720 sequences used in the analysis. Figure 5.7 depicts how often a sequence of a specific subtype (marked by the row label) was predicted as of another subtype (indicated by the column label). In general, the results were satisfactory. Subtypes B, C, D, AE, and O were predicted correctly almost always. The prediction engine also showed good performance on samples from subtype A or the recombinant AG. However, the results on the subtypes F and G left something to be desired. About 50% of the samples belonging to subtype F were predicted to be of subtype B. This closeness between the two clades is also reflected by the dendrogam depicted left to the image in Figure 5.7. With regards to subtype G we did not observe such a clear similarity to another subtype. Instead, the respective sequences were predicted to belong to the subtypes A/AG, B, or C.

We performed the same analysis also with predictions based on BLAST searches and multi-class SVMs. In all cases, the prediction models were trained on the same set of sequences and evaluated on the same test sets. In general, using the subtype of the top

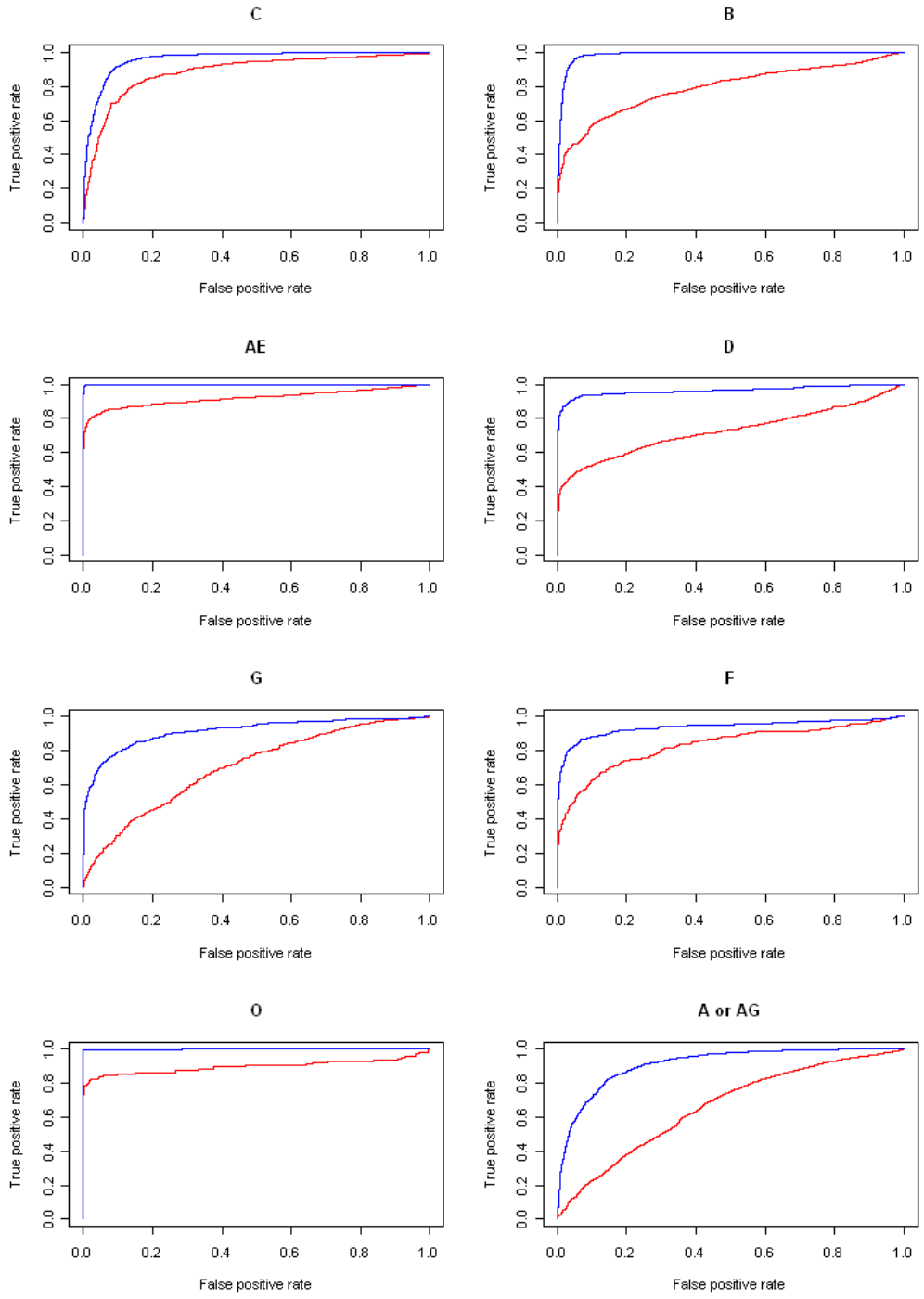


Figure 5.6: ROC curves for the Support Vector Machine models generated to predict the V3-subtype. The red curves display the predictions by the one-class SVM models whereas the blue curves show the respective prediction results of the two-class SVMs.

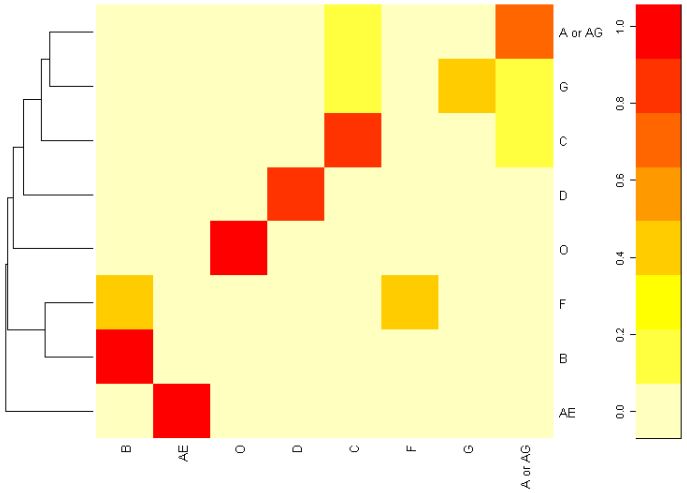


Figure 5.7: The figure shows how frequently a sequence of the subtype given by the row label was predicted to be of the subtype identified by the column label. The redder a box is the more often this combination occurred.

hit of a Blast search (*BLAST-1*) yielded the worst result. The average accuracy over ten runs was just 81.28%. Using the subtype appearing most often in the ten best results of the BLAST search (*BLAST-10*) lifted the average accuracy to 85.02%. Last but not least, when considering the subtype most often appearing in the 100 best hits of the BLAST output (*BLAST-100*), the accuracy slightly decreased again to 84.71%.

On the other hand, when predicting the V3-subtype with *MULTI* the average accuracy was slightly higher (87.5%) than using *TWO-STEP*. Nevertheless, we implemented *TWO-STEP* on the webserver GENO2PHENO[CORECEPTOR] because the improvement of *MULTI* over the use of *TWO-STEP* was not significant ($p = 0.139$ with Student's T-test) in terms of accuracy and because the individual prediction probabilities of the two-class SVMs can be used as features for coreceptor usage prediction.

**Performance of GENO2PHENO[CORECEPTOR] on B and non-B samples** In the next step we analyzed 738 clinically derived samples screened for maraviroc administration in Germany. Subtypes of the V3 loop sequences were predicted with our subtype predictor (*TWO-STEP*) and with the REGA subtyping tool. We first compared the predictions of the two tools. Due to the low number of non-B samples, we only compared subtype B against non-B predictions. Table 5.6 shows the numbers of predicted B and non-B isolates as well as the number of samples with no result. In the case of GENO2PHENO[CORECEPTOR], the subtype could not be predicted in two cases because of stop codons within the V3 loop sequence. Unfortunately, the REGA tool did not give a result in almost half of the sequences (49.7%), mainly because the bootstrap values were not good enough. However, the overall frequen-

cies of predicted B and non-B samples were very similar (geno2pheno[coreceptor]: 12.4% non-B vs. 15.4% non-B prediction by REGA). The agreement between the two tools was 96.2% on the samples for which both tools provided a result.

| *tool* | *B* | *non-B* | *no result* |
|---|---|---|---|
| geno2pheno[coreceptor] | 644 | 92 | 2 |
| REGA | 314 | 57 | 367 |

Table 5.6: Subtype predictions by geno2pheno[coreceptor] and REGA on the set of 738 V3 loops of clinically derived samples.

We next performed a coreceptor usage prediction with geno2pheno[coreceptor] and compared the predictions between predicted B and non-B samples. As can be seen in Figure 5.8, the performance on isolates predicted to be of subtype B is worse than coreceptor usage prediction on samples that were predicted not to be of subtype B. Using the 10% false-positive rate of geno2pheno[coreceptor], the sensitivity of B-samples was 60.3% at a specificity of 90.3%. At the same FPR-level, non-B viruses displayed a sensitivity of 85.7% and a specificity of 87.3%.

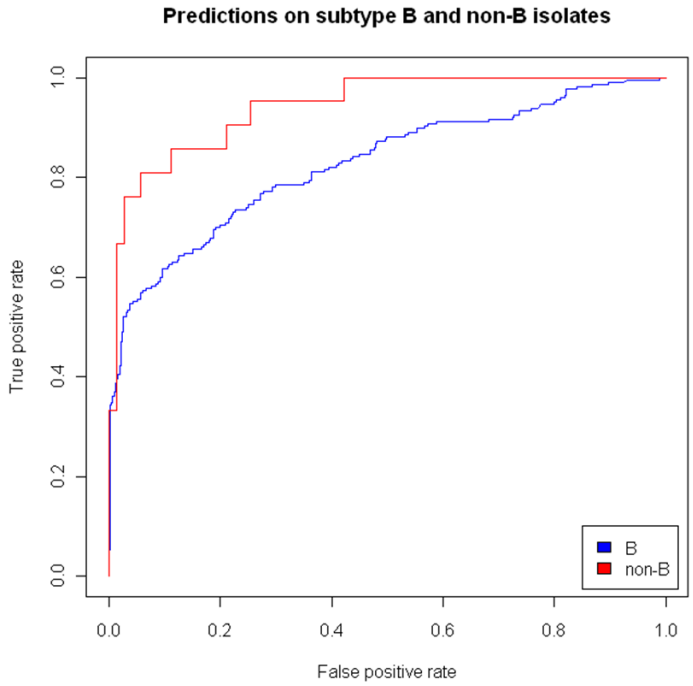


Figure 5.8: Coreceptor usage prediction on B and non-B isolates determined by geno2pheno[coreceptor]. The red curve shows the predictions from the set of non-B isolates whereas the performance on subtype B sequences is shown in blue. Subtypes were inferred from V3-genotype using the geno2pheno[coreceptor] subtype predictor.

### 5.3.3 Conclusions

In this section, we have shown how the V3-subtype can be determined from the V3 loop. In general, our results were quite encouraging because most subtypes could be predicted reliably. A reason why subtype G and subtype F predictions were not as good as the for the other subtypes might be the low number of sequences in the respective training sets. In both cases, we had less than 500 samples for training and validation. A larger dataset for training would probably help to improve the predictions significantly.

Our results were also in very good agreement with one of the most widely-used subtype prediction tools when tested on clinically derived samples. It should be pointed out that the high failure rate of the REGA subtyping tool is not a marker of inferiority in comparison to our tool but is due to the fact that REGA is optimized for working on *pol*-sequences rather than such short regions like the V3 loop. This was also one of the reasons why we decided to implement such a subtype prediction tool.

We do not want to claim that the subtype of the virus can be determined reliably by the V3 loop alone. In our opinion, the region is simply too short for a good prediction. Nevertheless, we find that it is still better to have an indication for a subtype than nothing. What we had in mind by the definition of this V3-subtype was to give the users a feeling of what to expect. If the predicted subtype were of subtype D which is known to be more often X4, then also the probability of being X4 should be higher. In contrast, if it is predicted to be A, then the tropism should be more probably R5.

As already mentioned in the results section we also tested other methods for subtype prediction. The predictions are clearly not perfect and several subtypes are still missing. We also mentioned that the multi-class SVM models seemed to outperform our two step approach but that we did not implement them in the current GENO2PHENO[CORECEPTOR] system. On the one hand, this was due to the fact that the implemented approach was easier to code and runs faster. On the other hand, we wanted to use the probabilities of the individual subtype predictions as features for coreceptor usage prediction. These features, however, did not yield any significant improvements in coreceptor usage prediction models, so far (data not shown). Thus, in the future one could consider exchanging the subtype predictor of GENO2PHENO[CORECEPTOR] and use multi-class SVMs instead.

When analyzing coreceptor usage prediction on the two subsets of samples, we encountered surprisingly good results on non-B isolates. On one hand, one might state that this outcome reflects the hypothesis that mutations defining coreceptor usage should be shared by the different subtypes. One the other hand, one could also argue that because the subtype predictor and the coreceptor usage prediction work relatively similar this might be a side effect. In fact, from our own observations we know that sequences looking *strangely* (not like usual V3 loops due to uncommon mutations or variations in the length of the loop) are more often predicted to be X4. Similarly, because they do not *look* like subtype B viruses they are probably more often been predicted to be non-B. However, while this could explain the high sensitivity of prediction it would not explain a high specificity. True or not, independent of the real correctness of the subtype predictions, we think that one can use the predicted subtype as a measure of confidence. For a sequence predicted to be not of subtype B, one can be more confident in the coreceptor usage prediction than for a predicted B virus which is in fact quite contradicting to the general opinion.

## 5.4 Clonal Vs. Clinical Data - is There a Difference?

Prediction of HIV-1 coreceptor usage dates back to the early 90's of the last century (e.g. Fouchier et al., 1992). Over the years, several prediction methods have been developed ranging from very simple rules to highly sophisticated machine learning methods. When tested on data from the freely available Los Alamos HIV Sequence Database, these methods generally perform very well (see Sing et al., 2007). However, when the first clinical datasets became available people were shocked to see the disappointing performance on these isolates (Low et al., 2007). In this thesis, we already used different approaches to elucidate why we encounter these dramatic differences between these two types of datasets. Within this section, we want to analyze if the sequences themselves look differently. This might be due to the fact that clonal data available in the database comes from older viral strains whereas samples collected in the clinics may display the current population which could have evolved from older variants over time. In fact, the virus might not only evolve within the host and adapt to the immune system but also the whole viral population might gradually evolve over time based on a natural adaptation to its host or due to antiretroviral drug treatment in the human population (see e.g. Ariën et al., 2007).

Also the stage of infection might play a role. We know that in the beginning of the infection R5-viruses dominate whereas over time more and more X4-viruses become detectable (Regoes and Bonhoeffer, 2005). A possible albeit not necessarily correct explanation might be that the immune system can control the X4-viruses at least to some extent. Over time, however, the virus evolves and the weakened immune system loses control more and more. Samples taken in the past and deposited in the Los Alamos HIV Sequence Database might have been collected from early and late stage infections similarly often as today. However nowadays, thanks to highly active antiretroviral therapy, people can live much longer with the virus which in consequence has more time to evolve.

Such ideas are of course just speculations but if there is some truth behind them we should see differences in sequences in the various datasets we have. Thus, in this part of the work we will compare data from the Los Alamos HIV Sequence Database, from a dataset containing therapy-naïve samples as well as datasets collected from therapy-experienced patients.

### 5.4.1 Material & Methods

**Datasets** Five different datasets were used in this analysis:

The *GERMANY* dataset comprises 864 samples from patients screened for maraviroc administration in Germany. The samples have been sequenced in several labs in Germany and the phenotype has been determined with both the original or the enhanced sensitivity Trofile assay (ESTA). 561 (64.9%) samples were phenotyped with either of the assays to be R5 whereas 303 (35.0%) samples had an X4-phenotype.

The second dataset (*ITALY*) was provided by the University Tor Vergata in Rome and consists of 465 samples screened for therapy with coreceptor antagonists in Italy. 320 samples (68.8%) were determined to be R5 and 145 samples (31.2%) as X4 by the Trofile assays (original and ESTA).

The *HOMER* and *MOTIVATE* datasets were both sequenced by the Center for Excel-

lence in HIV/AIDS in Vancouver, Canada. The HOMER dataset consists of 951 samples, 784 (82.4%) phenotyped of being R5 and 167 (17.5%) as X4 by a prototype of the original Trofile assay. All samples were from therapy-naïve patients. The MOTIVATE dataset contains 1913 samples from the screening population of the maraviroc phase III clinical studies. The samples were therapy-experienced and tropism determined with the Monogram Trofile Tropism Assay. 1213 (63.4%) of the samples were screened as R5 while 700 (36.6%) had an X4-phenotype.

Last but not least, all available V3-sequences with a phenotype were downloaded from the Los Alamos HIV Sequence Database (*LOSALAMOS*-dataset (L.A.)). These sequences are mainly clonal and have been phenotyped with different assays. At most one randomly chosen R5- and one X4-sequence per patient were included in the analysis. The dataset consists of 975 (82.1%) R5-viruses and 212 (17.7%) X4-viruses.

All samples of the different datasets were sequenced with conventional bulk-sequencing methods.

**Comparison of individual residues** For every position of the V3 loop and for every amino acid (including a gap for a possible deletion) we computed its frequency in (i) all samples, (ii) the R5-samples, and (iii) the X4-samples. The aforementioned comparison was performed for (i) all positions and residues in the alignment, (ii) amino acids significantly different ($p < 0.05$) between R5- and X4-samples in one of the five datasets, and (iii) residues significantly associated with X4-viruses in one of the different datasets. We then compared these differences between the different datasets using Pearson's correlation coefficient and displayed the results in scatter plots. Significance for differences between R5- and X4-samples were assessed using Fisher's exact test.

**Comparison of other properties** In addition to the comparison of the specific amino acids at different positions, we also tested for differences in other sequence features between the different datasets. We compared the prevalences of N-glycosylation sites in the V3 loop, the length and the charge of the loop, as well as the Livingstone-features: the number of hydrophobic, polar, small, tiny, aliphatic, aromatic, positively charged, negatively charged, and charged residues, and the number of prolines (Livingstone and Barton, 1993).

To account for differences in the prevalence of X4-viruses between datasets, we also tested for differences between R5-sequences and X4-sequences individually. Wilcoxon's rank sum test was used to assess statistical significance. Two datasets were reported to be different in a specific feature if there was a significant difference between their R5- as well as their X4-sequences. Significance was determined at different thresholds: $p = 5 \cdot 10^{-2}$, $p = 5 \cdot 10^{-3}$, $p = 5 \cdot 10^{-4}$, and $p = 5 \cdot 10^{-5}$.

### 5.4.2 Results

**Differences between amino acids frequencies among datasets** Figure 5.9 shows the frequencies of the individual amino acids in one dataset plotted against the respective frequencies in another dataset. In general, there was a very high correlation between individual datasets. The Pearson correlation coefficient was above 0.97 for all comparisons. However, while all comparisons between clinically derived samples were higher than 0.994, the

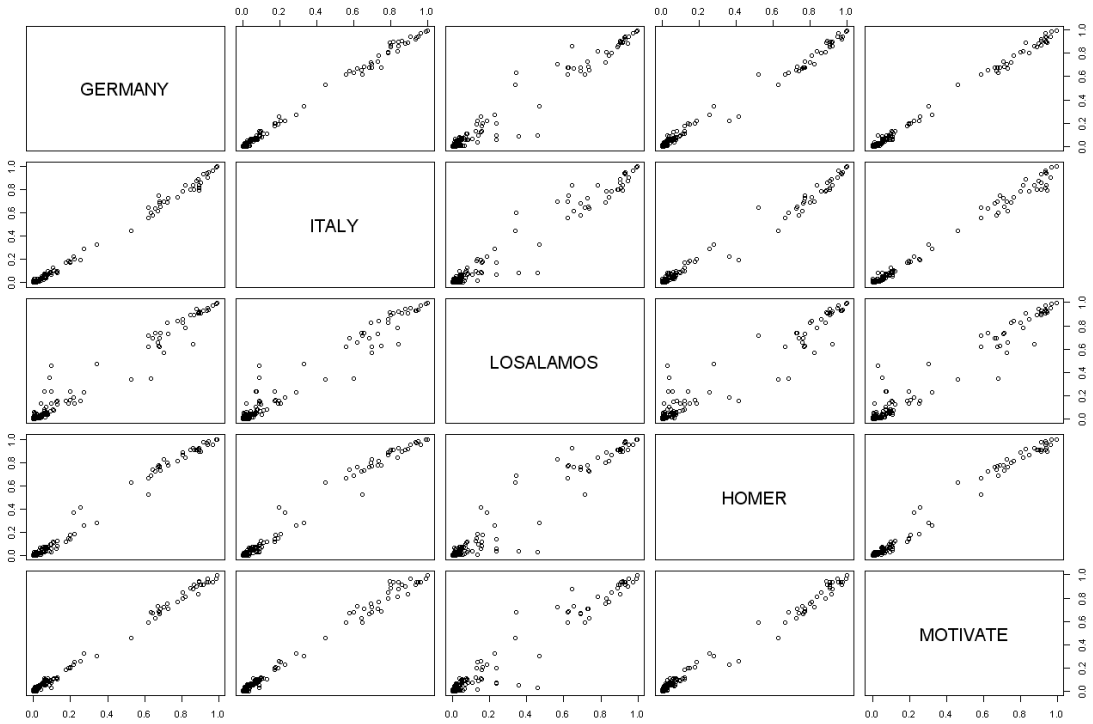


Figure 5.9: Frequencies of individual amino acids in one dataset plotted against the respective frequencies in another dataset.

highest value for a clinical dataset compared with the Los Alamos dataset was 0.986.

To account for differences in the prevalences of the two phenotypes between different datasets, we performed the same type of comparison for R5-viruses and X4-viruses individually. The results remained very similar: The correlation between Los Alamos samples and a clinical dataset was at most 0.978 when comparing frequencies in X4-samples and at most 0.987 when analyzing R5-viruses. When focusing only on the datasets comprising clinically derived samples, the correlation coefficient was at least 0.991 in all combinations (see Table 5.7).

Differences between the amino acid frequencies do not necessarily have to lead to problems in prediction systems. Mutations that are not predictive for the phenotype will usually not receive a high weight in the respective prediction models and thus will not change the predictions substantially. Therefore, we further focused on predictive amino acids. Only residues significantly associated with either R5- or X4-viruses in one of the five datasets were used in the next analysis.

The trends in the correlations were more or less the same as observed before. Neither when comparing all samples, nor when analyzing only R5- or only X4-viruses did any combination between the clonal dataset and a clinical dataset reach above 0.986. On the other hand, the lowest correlation for a combination among clinical datasets was 0.990 when X4-samples from Italy were compared with X4-samples from the MOTIVATE dataset.

Finally, we only compared the frequencies among amino acids that were predictive for X4-viruses. Table 5.8 lists the individual correlations between different datasets. The

| | ***GERMANY*** | ***ITALY*** | ***L.A.*** | ***HOMER*** | ***MOTIVATE*** |
|---|---|---|---|---|---|
| Differences between R5-viruses: | | | | | |
| ***GERMANY*** | X | 0.998 | 0.987 | 0.996 | 0.998 |
| ***ITALY*** | 0.998 | X | 0.986 | 0.994 | 0.997 |
| ***L.A.*** | 0.987 | 0.986 | X | 0.976 | 0.982 |
| ***HOMER*** | 0.996 | 0.994 | 0.976 | X | 0.996 |
| ***MOTIVATE*** | 0.998 | 0.997 | 0.982 | 0.996 | X |
| Differences between X4-viruses: | | | | | |
| ***GERMANY*** | X | 0.994 | 0.978 | 0.994 | 0.996 |
| ***ITALY*** | 0.994 | X | 0.974 | 0.994 | 0.991 |
| ***L.A.*** | 0.978 | 0.974 | X | 0.970 | 0.978 |
| ***HOMER*** | 0.994 | 0.994 | 0.970 | X | 0.993 |
| ***MOTIVATE*** | 0.996 | 0.991 | 0.978 | 0.993 | X |

Table 5.7: Correlations between the amino acid frequencies of different datasets in R5- and X4-viruses.

differences between clonal data and clinical data were even more obvious than before: While in all comparisons between clinical datasets a correlation of at least 0.982 was achieved, the correlations between the Los Alamos dataset and any clinical dataset never exceeded 0.930. The results are depicted in Figure 5.10 which illustrates that several amino acids predictive for X4-viruses occurred more often in clonal X4-viruses than in the respective viruses obtained in the clinics.

**Differences with respect to other properties** Coreceptor usage prediction is often not only based on specific mutations but also on features that do not rely on specific positions. For example, sometimes the overall charge of the V3 loop is used to discriminate between R5- and X4-viruses, or the presence of mutations at the N-glycosylation site in the V3 loop. Thus, we also compared the datasets with respect to such features.

Overall, we found 46 combinations of datasets in which the R5-viruses differed significantly ($p < 0.05$) in a property between the two datasets as well as the respective X4-viruses (see Appendix Table 2). We required that both R5- and X4-viruses had to be significantly different between datasets in order to account for differences in the tropism distribution. This was done because if we pooled R5- and X4-sequences and a feature like e.g. the charge of the V3 loop were significantly different between R5- and X4-samples, then we would of course expect to see the feature to be different when comparing datasets with different X4-prevalences, too. Consequently, we would introduce bias into our analysis.

In seven comparisons, the number of hydrophobic residues was significantly different in two datasets at a significance threshold of 5%. Additionally, the number of small (6 times), tiny (5 times) and positively charged (5 times) residues as well as the overall charge (5 times) were different between two datasets relatively often.

When lowering the significance threshold it became obvious that especially the number of tiny residues often differed significantly between datasets (4 times at a significance value of $p = 5 \cdot 10^{-6}$). These four combinations were the comparisons between the Los Alamos dataset and either of the clinical datasets in which in all cases the number of tiny residues

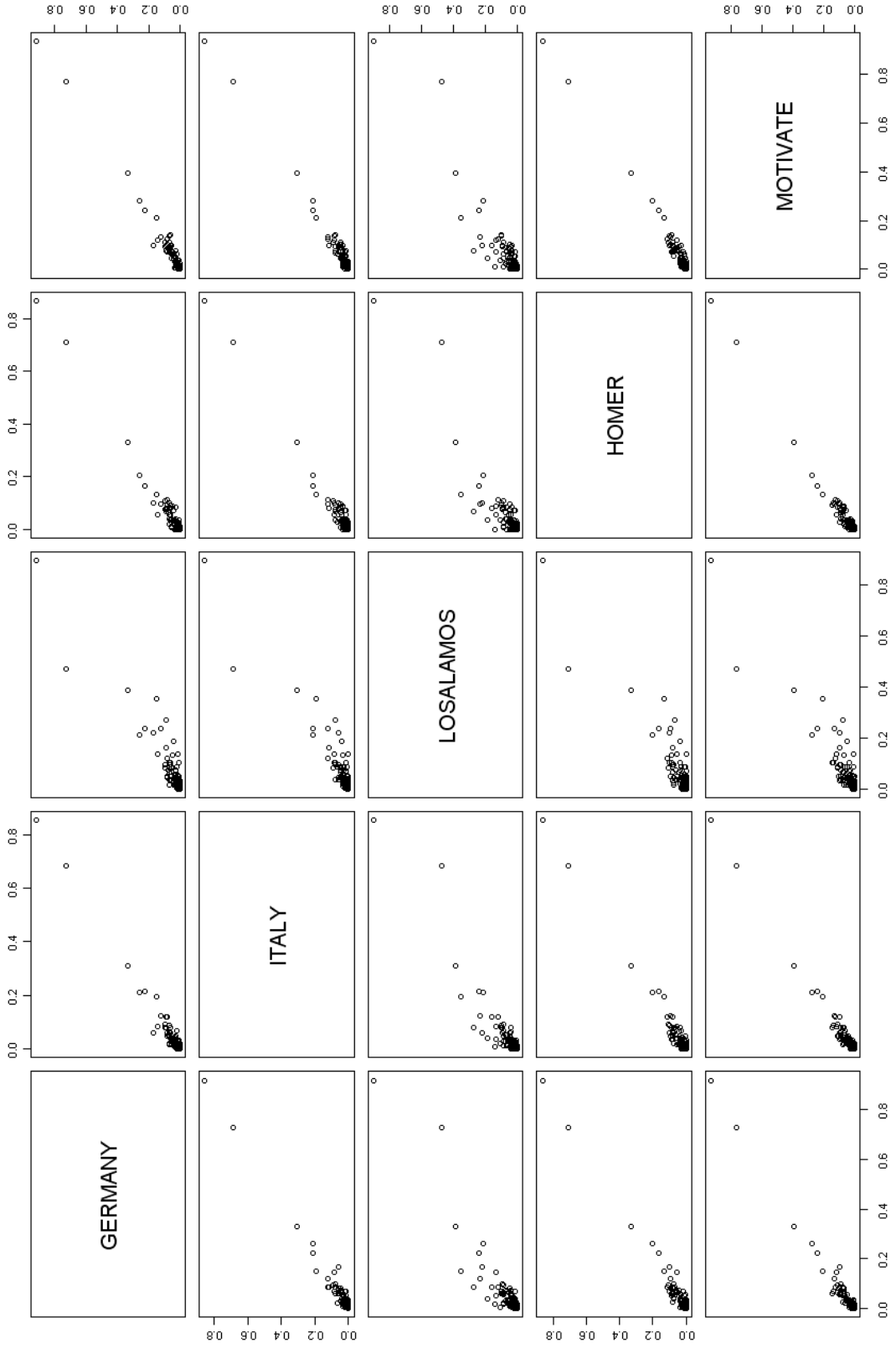


Figure 5.10: Frequencies of amino acids predictive for X4-viruses in one dataset plotted against the respective frequencies in another dataset.

| | ***GERMANY*** | ***ITALY*** | ***L.A.*** | ***HOMER*** | ***MOTIVATE*** |
|---|---|---|---|---|---|
| Differences between all viruses: | | | | | |
| ***GERMANY*** | X | 0.996 | 0.925 | 0.992 | 0.994 |
| ***ITALY*** | 0.996 | X | 0.930 | 0.991 | 0.993 |
| ***L.A.*** | 0.925 | 0.930 | X | 0.904 | 0.903 |
| ***HOMER*** | 0.992 | 0.991 | 0.904 | X | 0.988 |
| ***MOTIVATE*** | 0.994 | 0.993 | 0.903 | 0.988 | X |
| Differences between R5-viruses: | | | | | |
| ***GERMANY*** | X | 0.996 | 0.920 | 0.992 | 0.995 |
| ***ITALY*** | 0.996 | X | 0.927 | 0.989 | 0.993 |
| ***L.A.*** | 0.920 | 0.927 | X | 0.892 | 0.893 |
| ***HOMER*** | 0.992 | 0.989 | 0.892 | X | 0.991 |
| ***MOTIVATE*** | 0.995 | 0.993 | 0.893 | 0.991 | X |
| Differences between X4-viruses: | | | | | |
| ***GERMANY*** | X | 0.984 | 0.917 | 0.984 | 0.987 |
| ***ITALY*** | 0.984 | X | 0.922 | 0.982 | 0.987 |
| ***L.A.*** | 0.917 | 0.922 | X | 0.904 | 0.917 |
| ***HOMER*** | 0.984 | 0.982 | 0.904 | X | 0.987 |
| ***MOTIVATE*** | 0.987 | 0.987 | 0.917 | 0.987 | X |

Table 5.8: Correlation between different datasets in the frequencies of amino acids significantly associated with X4-viruses.

was significantly lower in the clonal dataset than in the respective clinical dataset in both R5- and X4-samples. Thus, if the Los Alamos dataset were used for training of a coreceptor usage prediction system, this feature should not be used or should at least receive a lower weight.

The Los Alamos dataset was also involved in the other four combinations (twice the number of hydrophobic residues, once the number of small and negative residues, respectively) of datasets in which a feature was significantly different at a significance level of $p = 5 \cdot 10^{-6}$.

Overall, in most combinations in which a feature differed between two datasets the Los Alamos dataset was one of them. The lower the significance threshold was chosen the more often the difference was between one of the clinical datasets and Los Alamos data. Since ten different combinations were tested per feature with a dataset being involved in four of these ten, one would expect each dataset to be involved in 40% of the significant cases. However, Table 5.9 clearly shows that the Los Alamos dataset appeared more often than expected (with $p < 0.001$ at all significance levels using a one-sided binomial test), especially the lower the significant threshold was selected.

### 5.4.3 Discussion

Overall we have observed very high correlations between the different datasets suggesting that there is no major bias between most of them irrespective of the stage of infection.

We did not see a clear difference between isolates from therapy-naïve and therapy-

| | $p = 5 \cdot 10^{-2}$ | $p = 5 \cdot 10^{-3}$ | $p = 5 \cdot 10^{-3}$ | $p = 5 \cdot 10^{-4}$ | $p = 5 \cdot 10^{-5}$ |
|---|---|---|---|---|---|
| ***#sig. comb.*** | 46 | 29 | 18 | 11 | 8 |
| ***GERMANY*** | 30.4% | 17.2% | 11.1% | 18.2% | 12.5% |
| ***ITALY*** | 19.6% | 13.8% | 5.6% | 9.1% | 12.5% |
| ***LOSALAMOS*** | 65.2% | 75.9% | 88.9% | 90.9% | 100.0% |
| ***HOMER*** | 41.3% | 44.8% | 44.4% | 45.5% | 37.5% |
| ***MOTIVATE*** | 43.5% | 48.3% | 50.0% | 36.4% | 37.5% |

Table 5.9: Amount of features significantly different between two datasets at different significance thresholds and how often the respective datasets were involved.

experienced patients so that one can assume that the prediction models should work equally well on these individuals. However, in all of our analyses data from the Los Alamos Database showed a lower correlation with the clinical datasets in comparison to the correlations comparing the clinical datasets with each other. This is of course unfortunate as all current prediction systems have been trained on this type of data.

One might argue that minor variants not detected by standard population-based sequencing in clinical isolates could be the reason for the shifts in frequencies of certain residues in X4-viruses between clonal clinical data (see also Chapter 4). To some extent this is certainly true but it does not explain why we see similar differences in the R5-viruses and why the correlations between the clinical datasets are much higher.

Even if the shifts in frequencies would be due to minor variants, these would still have a negative impact on our predictions because it would overemphasize some mutations that are usually not found in the viral quasispecies. A further point arguing against this theory is that it seems that specific residues always differed between the Los Alamos dataset and the clinical datasets. Because these are observed among all clinical datasets, it is unlikely that they arise from undetected minor populations.

It is surprising, how well the amino acid frequencies correlate among the clinical datasets. Especially, when comparing sequences from therapy-naïve patients of the HOMER cohort with samples from the heavily pre-treated study population of MOTIVATE, one might have expected more deviations. It should also be pointed out here that the sequences of the clinical datasets are derived from different labs which use different sequencing protocols and different groups of patients. Therefore, some natural variation would have been expected but the correlations we have seen were impressively high.

Similar results were found when position-independent properties of the V3 loop were analyzed. Again, most differences between two datasets occurred when one of these datasets was the Los Alamos HIV Sequence Database. While most of the tested features are not currently implemented in coreceptor usage prediction systems, some of them like the charge or the presence of mutations in the N-glycosylation motif are and should therefore been inspected carefully in future.

At first sight, one might think of using the clinical datasets for training the prediction models rather than clonal data. However, this would also have several drawbacks: First of all, we know that the Trofile assay is more sensitive than standard population sequencing methods (see also Chapter 4), i.e. it might happen that in a viral quasispecies containing a minority of X4-viruses, these would not have been detected by population sequencing.

Instead, one would sequence the major variant possibly being R5 rather than the X4-viruses the phenotype detected. Thus, we would induce a bias into our training set by labeling these R5-sequences as X4. In case of clonal data on the other hand, we can be relatively sure that the sequence expresses the phenotype.

Another factor inducing these differences might be the distribution of subtypes. In the Los Alamos database several samples are from different subtypes whereas the clinical datasets used in this analysis are mainly comprised of subtype B samples, especially the HOMER (comprising almost only subtype B viruses) and the MOTIVATE dataset. Training on these datasets might therefore impair the predictiveness on other subtypes.

Such problems with using clinically derived samples for training of HIV-1 coreceptor usage prediction tools might perhaps be solvable by sophisticated methods like covariate shift adaptation techniques. Covariate shift models allow for differences between training and test distribution of instances and have already been applied successfully in the realm of HIV therapy selection (Bickel et al., 2008).

Altogether, we therefore conclude that there are differences between clonal and clinical datasets and that one should be aware of this problem when training prediction models. However, at present we do not see a good way to replace the clonal dataset with clinically derived samples or to lower the weights for the differing residues in a systematic way.

## 5.5 Coreceptor Usage Prediction from Bulk-Genotype for Therapy-Selection

Prediction of HIV-1 coreceptor usage from genotype is an interesting research topic. In the previous sections and chapters we have shown how to improve these predictions. However, the main goal of our work is to help people infected with HIV by selecting appropriate therapies for them.

With the advent of HIV-1 coreceptor antagonists the topic of HIV-1 tropism came into focus of daily diagnostics, too. Naturally, a coreceptor antagonist works best, if the virus only can use the receptor this drug blocks. And with a phenotypic coreceptor test we can measure this phenotype quite well. We have seen that genotypic approaches to determine tropism seem to be less sensitive than phenotypic tests but if the comparison assay is the ground truth of course one cannot achieve perfect results. Our real aim is therefore not to mimic the results obtained by phenotypic assays like the Monogram Trofile Tropism Assay but to predict if patients will respond to therapies containing coreceptor antagonists or not.

CCR5 antagonists are a promising drug class and have recently been added to the arsenal of anti-HIV drugs. The first compound to be approved by the FDA and the EMEA was maraviroc (MVC, Celsentri) from Pfizer (Dorr et al., 2005). This drug is currently administered for treatment of therapy-experienced patients in Europe and North America but to treatment-naïve patients only in the United States (Gulick et al., 2008; Lieberman-Blum et al., 2008; Cooper et al., 2010). Because MVC blocks only the CCR5-receptor, viruses capable of using CXCR4 are not susceptible to it. Thus, a tropism test is mandatory before administration of the drug.

Approval for therapy-experienced patients was based on the clinical trials MOTIVATE-1 and MOTIVATE-2, as well as the safety study A4001029 (1029-study). In the MOTIVATE studies, the efficacy of the drug was compared against placebo in combination with optimized background therapy (OBT) in patients screened who harbored only R5-viruses. The studies were double-blinded and patients were randomized in a 1:2:2 ratio to receive either placebo (PCBO-arm), once daily dosis of maraviroc (QD-arm), or twice daily dosis (BID-arm). Approximately 45% of patients receiving MVC had undetectable HIV-1 RNA levels after 48 weeks of medication (Gulick et al., 2008).

The 1029-study enrolled patients screened out of the MOTIVATE-studies due to non-R5 using viruses or for which no phenotype could be determined. This study was performed as a safety study to show that the administration of MVC does not harm patients carrying an undetected X4-virus. This concern was raised because CXCR4-using viruses predominantly appear in late stage disease and are associated with faster disease progression. As expected, MVC did not significantly reduce viral load in comparison to placebo (Saag et al., 2009).

Determination of HIV-1 coreceptor usage for the clinical trials was performed using the original phenotypic Monogram Trofile Tropism Assay (Monogram Biosciences, South San Francisco) (Whitcomb et al., 2007). In this section, we want to retrospectively analyze this data and assess the quality of genotypic coreceptor usage prediction from conventional bulk-sequenced isolates with respect to treatment outcome. In this way we want to understand how good or bad the genotypic approach really is when used in daily routine.

The work of this section has been published in (McGovern et al., 2010) and presented

on several conferences. It was carried out in a joint work by collaborators from the British Columbia Center for Excellence in HIV (Vancouver, British Columbia, Canada), Pfizer Inc. (New York, USA), and Pfizer Research & Development (Sandwich, UK). Because the predictions should be carried out fully blinded without prior knowledge of clinical outcome our part in this analysis was mainly restricted to generating the genotypic prediction scores.

### 5.5.1 Materials & Methods

**Study Population** Patient isolates were taken from the MOTIVATE-1, MOTIVATE-2 and 1029 studies. Subjects enrolled into the MOTIVATE-1 and -2 studies required to have a minimum plasma RNA viral load of 5000 copies per milliliter blood and therapy-experience (Gulick et al., 2008). In total, 3244 patients were screened for these studies and their tropism determined by the Trofile assay. 1049 of the screened patients entered one of the clinical trials that were performed in Canada and the United States (MOTIVATE-1) and Australia, Europe and the United States (MOTIVATE-2), respectively. Patients with an X4- or D/M-virus, as well as patients for which no phenotypic result could be generated were excluded from these studies (N=955). A subset of these patients screened to have "non-R5"-virus, were enrolled into the A4001029 study (N=186) (Saag et al., 2009). Because the different studies were of identical design data from all studies could be pooled for this analysis.

**Sequencing** Viral RNA was extracted from frozen plasma samples using the automated, standard operating procedures for the NucliSENS easyMAG (bioMérieux). Using nested RT-PCR methods, the V3 loop of gp120 from the HIV-1 env gene was amplified in triplicate from the RNA extracts. Reverse transcriptase (RT) and first-round PCR reactions were combined into one step by employing SuperScript™ III One-Step RT-PCR system with Platinum® Taq High Fidelity enzyme (Invitrogen) alongside forward primer SQV3F1 (5' GAG CCA ATT CCC ATA CAT TAT TGT 3') and reverse primer CO602 (5' GCC CAT AGT GCT TCC TGC TGC TCC CAA GAA CC 3'). Second-round PCR was performed using Expand™ High Fidelity enzyme (Roche Diagnostics) with forward primer SQV3F2 (5' TGT GCC CCA GCT GGT TTT GCG AT 3') and reverse primer CD4R (5' TAT AAT TCA CTT CTC CAA TTG TCC 3'). Second-round PCR amplicons were subsequently sequenced in both the 5' and 3' directions using the ABI 3730 automated sequencer (Applied Biosystems) using BigDye® Terminator (Applied Biosystems) with forward primer V3O2F (5' AAT GTC AGY ACA GTA CAA TGT ACA C 3') and reverse primer SQV3R1 (5' GAA AAA TTC CCT TCC ACA ATT AAA 3').

**Sequence Interpretation** Consensus sequences were generated for each sample using population sequencing. Electropherograms were analyzed with the fully automated custom software *ReCall* (Brooks et al., 2009). Coreceptor usage of the V3 loop sequences was predicted using GENO2PHENO[CORECEPTOR]. Predictions were performed fully blinded, i.e. no prior knowledge of clinical outcome was available. For each consensus sequences of the three PCR amplification products, a prediction was performed and the one with the lowest predicted false-positive rate was used. To account for the triplicates the false-positive rate cutoff was lowered and arbitrarily chosen to be 5% in order to stratify patients as either

R5 or non-R5 for clinical outcome analysis.

**Outcome measures and predictor variables** The median change in plasma viral load after 8 weeks of treatment was selected to be the primary endpoint for the analysis using a continuous outcome. Patients were classified to be responder if their viral load at week 8 was undetectable (<50 copies per milliliter) or if their logarithmic viral load decrease was at least 2 in comparison with baseline viral load. Last observation carried forward was performed in case of missing data at week 8. Secondary analysis was performed at week 24. Patients with missing data at this time point were considered to have failed treatment. Comparisons were done between patients infected with R5 and non-R5 virus and between genotypic and phenotypic determinations.

The impact of MVC independent of optimized background therapy (OBT) was assessed by using a weighted susceptibility score (wSS) derived from phenotypic susceptibility and drug class potency (Valdez et al., 2008). Data was stratified by OBT into three groups with wSS <1, 1 to <2, and ≥2. For this purpose, only the drugs in the treatment regimen at day 1 of therapy were considered. In case of phenotypic drug resistance or "inactive" agents, a score of 0 was given. NRTIs were given a score of 0.5 and both NNRTIs and PIs were given a score of 1. The median change in viral load for each wSS category and each screening method was compared between the three study arms.

### 5.5.2 Results

**Changes in median viral loads** Based on the similar study designs, samples from the approval studies MOTIVATE-1, MOTIVATE-2, and A4001029 were pooled and virologic outcomes were compared between the different screening methods. Changes in median viral load decreases in the different groups receiving placebo, once daily, and twice daily drug dosis were found to be highly comparable irrespective of the different approaches used.

The dataset contained 462 patients in the BID-arms of the different studies. Among these, the Trofile assay determined 405 (87.7%) patients to have only R5-virus and 57 (12.3%) to have CXCR4-using variants. geno2pheno[coreceptor] found 393 (85.1%) of the patients receiving maraviroc twice daily eligible for treatment while 69 (14.9%) were predicted to have X4-viruses. Patients screened with R5-viruses by either Trofile or geno2pheno[coreceptor] experienced a median decrease in viral load of 2.4 logs, significantly greater than that of patients screened with X4-viruses by the respective assays (see Figure 5.11).

The QD-arm revealed comparable results in terms of median change in viral load. Trofile determined 390 of the 444 patients (87.8%) whereas geno2pheno[coreceptor] found 374 patients to have R5-virus (84.2%). In the last study arm in which placebo was given instead of maraviroc, no difference was observed as expected. Viral load decreases were similar among different assays and tropism results. Both assays also found similar numbers of X4-viruses (Trofile: 58; geno2pheno[coreceptor]: 50) among the 258 patients in this treatment arm.

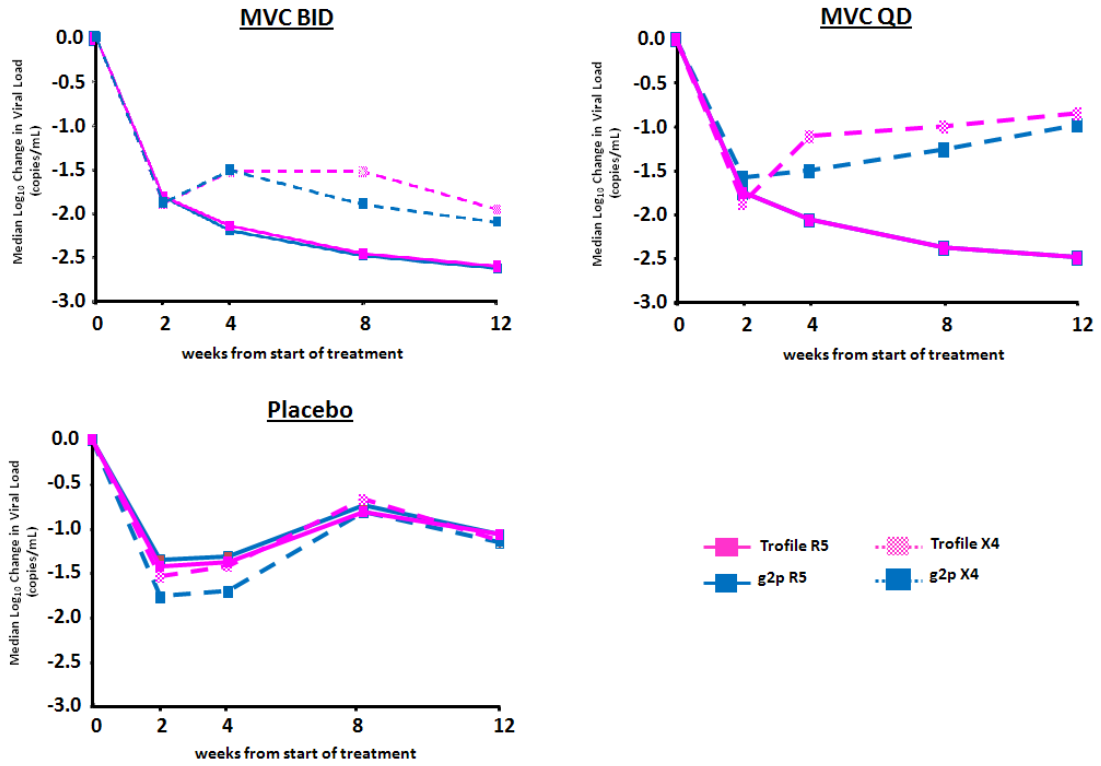


Figure 5.11: Virological response stratified by screening assays for patients who entered the MOTIVATE-1, MOTIVATE-2 and 1029-studies. Screening results obtained with Trofile are displayed in magenta whereas results by GENO2PHENO[CORECEPTOR] are shown in blue. Solid lines indicate median viral load changes for patients with R5 results, while dashed lines indicate non-R5 results. The three treatment arms were maraviroc BID (MVC BID), maraviroc QD (MVC QD) and Placebo.

**Comparison of predicted responders and non-responders** When comparing the percentage of patients responding to therapy, no significant difference between the assays could be observed at week 8. In the BID-arm, 272 (67.2%) of the 405 patients determined to have only R5-virus by Trofile had a virological response whereas 25 (43.8%) of the 57 non-R5 patients achieved viral suppression at week 8. When using GENO2PHENO[CORECEPTOR] to determine tropism, 263 (66.9%) of the 393 patients predicted to have R5-virus and 34 patients with X4-virus (50%) responded to therapy (see also Table 5.10). Again, similar results were found for the QD-arm and were also consistent independent of the definition of virological success (required reduction of 1, 1.5, 2.5, or 3 logs, data not shown). Not surprisingly, in the Placebo arm virological response was associated with tropism neither when determined with the Trofile assay nor when genotypic interpretation was used.

**Stratification according to background therapy** When using the weighted sensitivity score to stratify patients into three groups, no difference was found between Trofile and

| *Assay* | *Tropism* | *R* | *NR* | *RR* | *SE* | *SP* |
|---|---|---|---|---|---|---|
| A) BID-arm | | | | | | |
| Trofile | R5 | 272 | 133 | 67% | 92% | 20% |
| | non-R5 | 25 | 32 | 44% | 91% | 19% |
| geno2pheno[coreceptor] | R5 | 263 | 130 | 67% | 89% | 24% |
| | non-R5 | 34 | 35 | 49% | 89% | 21% |
| B) QD-arm | | | | | | |
| Trofile | R5 | 250 | 140 | 64% | 92% | 21% |
| | non-R5 | 23 | 31 | 43% | 92% | 18% |
| geno2pheno[coreceptor] | R5 | 240 | 133 | 64% | 88% | 25% |
| | non-R5 | 33 | 38 | 46% | 88% | 22% |

Table 5.10: Patients achieving virological response at week 8. R: responders, NR: non-responders, RR: response rate; SE: sensitivity; SP: specificity

geno2pheno[coreceptor] (see Figure 5.12). In total 316 patients had a weighted sensitivity score below 1. Of these 19 patients were phenotyped to be non-R5 by Trofile and 31 X4 by geno2pheno[coreceptor]. Viral load decreased in patients with determined R5-virus by 1 to 1.5 logs over the first 8 weeks of therapy in comparison to patients with predicted X4-virus. The difference in viral load decrease was even more profound when geno2pheno[coreceptor] was used to assess tropism than using Trofile.

For the other two groups no difference was found between the two assays. Patients with a wSS of 1 to $< 2$ (N = 347) benefited in terms of viral load decreases by 1 to 2 logs over the first 8 weeks of therapy when the phenotype was R5 (Trofile: 293 patients, geno2pheno[coreceptor]: 290 patients). In the group of patients with a highly active therapy backbone (wSS of $\geq 2$) tropism prediction did not seem to be informative. The median viral load decreased in this group for more than 2.5 logs regardless of predicted tropism and assay used to determine it.

**Stratification according to assay concordance** We also compared the decreases of median viral load for patients in which Trofile and geno2pheno[coreceptor] agreed or disagreed, respectively. Figure 5.13 depicts the change of median viral load in the first 12 weeks after starting therapy with maraviroc (once or twice daily dosis). The two assays agreed in 817 of 906 samples (90.2%). 737 had concordant R5-calls whereas 80 were determined to be X4 by both assays. The former group of patients had a median viral load decrease of 2.5 logs at week 8 whereas the viral load of the second group decreased by only 1.0 logs by week 8.

89 patients had discordant results. 31 (3.5%) of these were called non-R5 by Trofile but R5 by geno2pheno[coreceptor], while 58 patients (6.5%) were called R5 by Trofile but non-R5 by geno2pheno[coreceptor]. However, both groups showed similar median viral load changes of -2.0 log copies per milliliter after 8 weeks of treatment with maraviroc, regardless of the direction of discordance and were thereby placed in between concordant R5- and concordant X4-viruses.

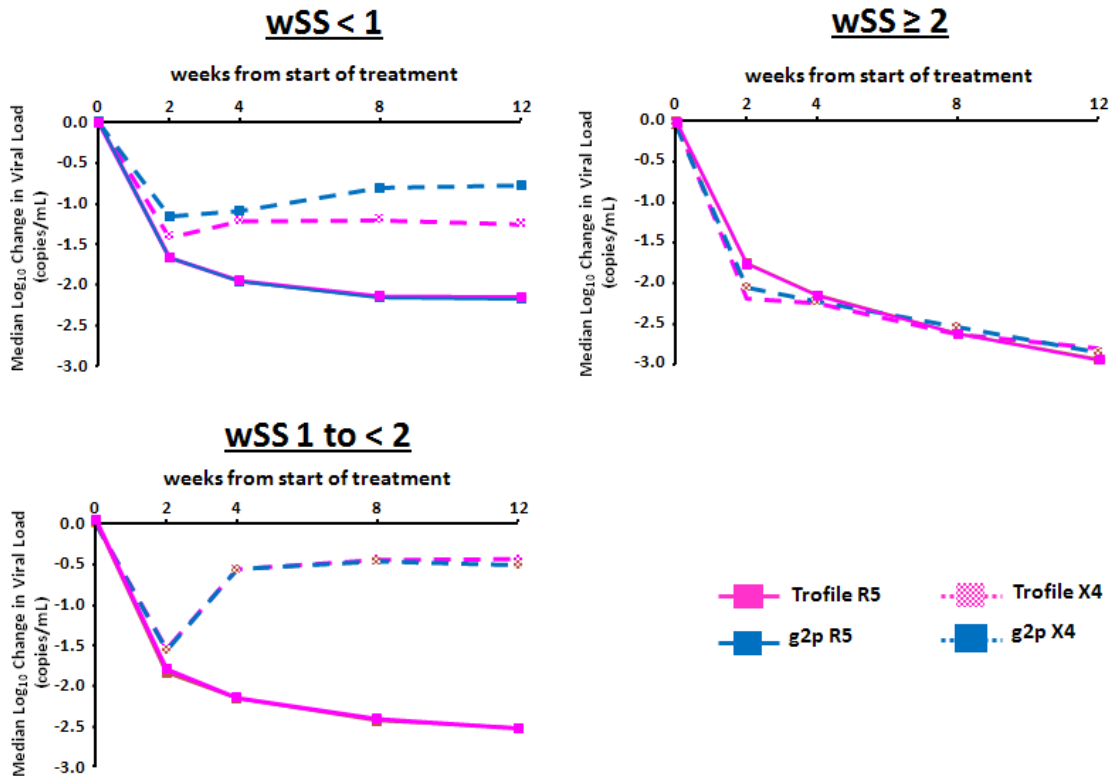


Figure 5.12: Virological response stratified by activity of optimized background therapy for patients who entered the MOTIVATE-1, MOTIVATE-2 and 1029-studies. Screening results obtained with Trofile are displayed in magenta while results by geno2pheno[coreceptor] are shown in blue. Solid lines indicate median viral load changes for patients with R5 results, while dashed lines indicate non-R5 results. The treatment arms maraviroc BID and maraviroc QD were combined for this analysis. The activity of the background regimen was assessed using a weighted sensitivity score (wSS) reflecting the sensitivity and inherent potency of the background regimen.

### 5.5.3 Conclusions

In this study we examined the predictive power of the Monogram Trofile Tropism Assay and geno2pheno[coreceptor] with respect to virological outcome. Previous studies comparing genotypic approaches to infer viral tropism with the Trofile assay showed poor results. This led to the conclusion that they would be "inadequate" for daily diagnostics (e.g. Low et al., 2007). This work, however, indicates that genotypic interpretation can be used to screen patients for administration with coreceptor antagonists. We have demonstrated that geno2pheno[coreceptor] provides broadly similar outcomes in virological response as the phenotypic test in a large dataset of patients receiving maraviroc. The results were consistent over 24 weeks of observation and across several subsets.

Tropism determination was especially predictive for therapy outcome in the groups of patients which had not a fully active background therapy (wSS $< 2$) whereas patients with

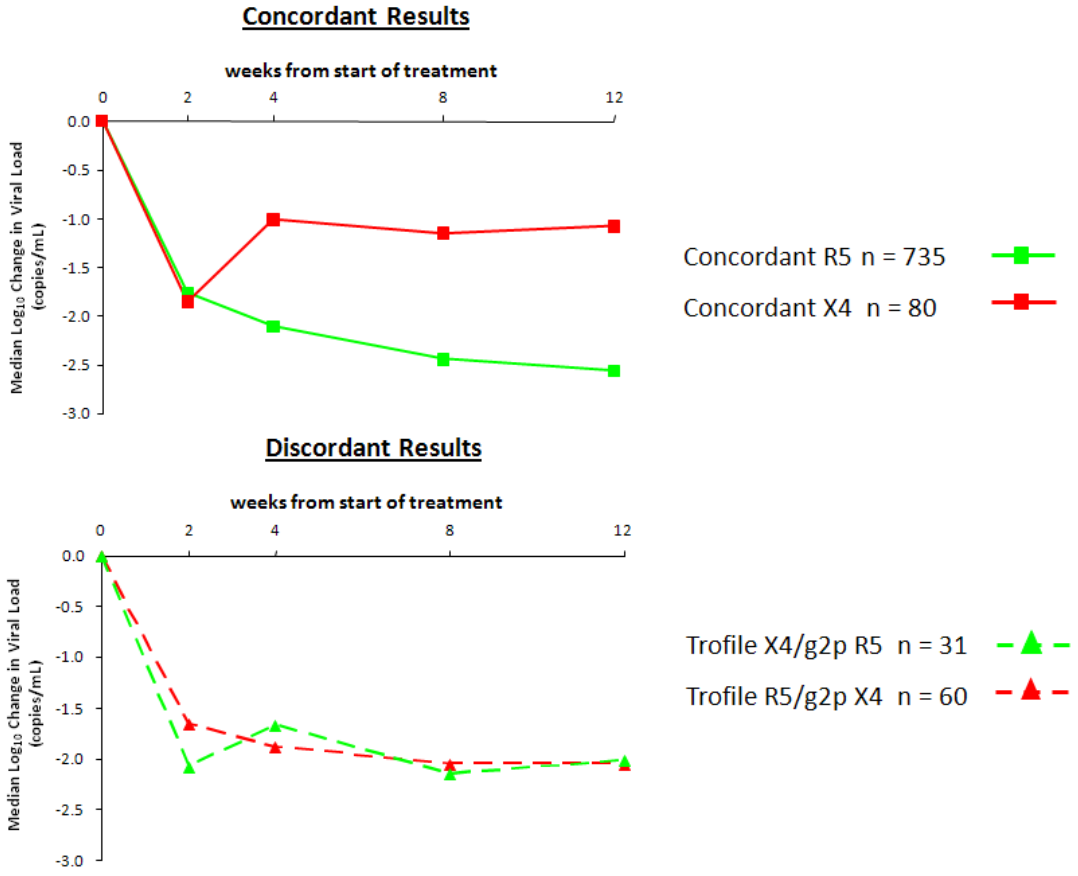


Figure 5.13: Virological response stratified by assay concordance for patients who entered the MOTIVATE-1, MOTIVATE-2 and 1029-studies. The treatment arms maraviroc BID and maraviroc QD were combined for this analysis. The upper panel indicates median viral load changes where both assays indicated R5-virus (green) and where both assays indicated non-R5-virus (red). In the lower chart, median viral load changes are depicted in which the two assays disagreed. red: R5 by Trofile but X4 by geno2pheno[coreceptor], green: X4 by Trofile but R5 by geno2pheno[coreceptor].

a highly active therapy backbone showed a median viral load decrease of more than 2.5 logs irrespective of determined tropism. This might indicate that two fully active drugs in a regimen are sufficient for early antiviral response but should not necessarily imply long term efficacy.

A fascinating property of the dataset used in this analysis is that also patients carrying CXCR4-using viruses were enrolled. Thus, also patients could be analyzed who responded to therapy despite of a negative result by the Monogram Trofile Tropism Assay. Such data is usually not available as it is considered unethical to give patients a drug which should not work in theory. In this study, however, it was possible because the safety of the drug was assessed. Furthermore, the study dataset is unique because the different treatment arms were randomized, multinational registrational trials allowing the investigation of large numbers of very well-characterized patients. Therefore, also the activity of the backbone

therapy could be analyzed in detail.

A notable difference to previous studies is that sequencing was performed in triplicates, i.e. using three independent RT-PCRs. This was done to increase the likelihood of identifying minority species because the detection limit for population-based sequencing is approximately 15-30% and highly dependent on the input copy number of the virus. In a related study in which we took part, we could analyze the same dataset using massively parallel sequencing (Swenson et al., 2010c). When compared to the results presented here, "deep" sequencing showed an increased ability to detect low abundance minority X4 species, though only a modest improvement in terms of predicting virological response following the start of MVC.

The phenotypic assay used to screen for patients eligible for maraviroc administration was the original Monogram Trofile Tropism Assay. Unfortunately, no phenotypic result from the enhanced sensitivity Trofile assay was available. However, as already shown in Section 5.2, this deficit might not be too problematic because one can adjust the FPR cutoff of geno2pheno[coreceptor]. Moreover, we recently also analyzed data from the MERIT-trial analyzing the efficacy of maraviroc in treatment-naïve patients. The drug did not reach the anticipated study goal and therefore has not been approved for treatment-naïve patients in Europe. However, when the samples were re-screened with the enhanced sensitivity Trofile assay, it could be shown that non-inferiority compared with efavirenz would have been achieved. Thus, the drug has now also been licensed by the FDA (Cooper et al., 2010). We could demonstrate that using population-based sequencing as well as deep-sequencing of the V3 loop and inferring coreceptor usage from genotype would have yielded comparable results as achieved by the ESTA assay.

Altogether, our results suggest therefore that genotypic interpretation of the V3 loop can be used to predict early antiviral response to maraviroc in daily routine. This allows for more rapid turnaround times, broader availability and lower costs.

# 6 Conclusions and Outlook

Coreceptor antagonists have become a valuable addition to the arsenal of antiretroviral drugs against HIV-1 in the last few years. Assessing viral tropism before administration of these drugs is essential and mainly based on phenotypic assays rather than genotypic approaches. Within this work we have tackled the problem of genotypic prediction of HIV-1 coreceptor usage from different angles.

Our first approach was to move the focus away from analyzing the V3 loop alone to the incorporation of other regions within and beyond the envelope gene. The V2 loop of gp120 was already known to be part of the bridging sheet forming the coreceptor binding site and mutations within this loop were also known to affect viral tropism. We could show that this property can also be used to improve HIV-1 coreceptor usage prediction both on clonal and on clinical data. Another region we focused on was the gp41-protein normally not related to coreceptor usage. However, recent studies indicated some associations with tropism and we could find several amino acid substitutions highly significant for phenotype. The discovery of an insertion of a hydrophobic residue at the N-terminal end of gp41 related to X4-viruses was surprisingly and its predictive power even reached that of the 11/25 rule. We could also demonstrate that gp41-sequence information can be used to reliably predict HIV-1 coreceptor usage in two large and well-defined cohorts. However, although being highly predictive, the incorporation of gp41 into a model using also the V3 loop as input did not improve prediction performance substantially neither when phenotype nor when clinical outcome was predicted. Last but not least, we analyzed other regions in the envelope protein and other viral proteins for correlation with tropism. Although some results were interesting and encourage further analysis, we do not expect major impact on the quality of HIV-1 coreceptor usage prediction methods.

The second main part of our work dealt with analyzing the viral quasispecies with next generation sequencing using the 454-technology. A common explanation for the impaired coreceptor usage prediction results on clinically derived data is the inability of standard Sanger sequencing to detect minor variants. Next generation sequencing capable of detecting minorities to very low levels was therefore a promising technology to improve predictions. However, several issues had to be resolved first. We have therefore shown how to process the data in an efficient way, how to detect and treat sequencing errors, and how to present the results in different manners. In a proof-of-concept study we then demonstrated how next generation sequencing can be combined with current coreceptor usage prediction methods. Our prediction results were concordant with the phenotypic assay in all but one case. Based on these results, we finally developed a new webserver predicting viral tropism from samples sequenced with the 454-technology.

Current prediction of HIV-1 coreceptor usage in daily routine is based neither on regions beyond V3 nor on using next generation sequencing technologies. The last part of our work therefore addressed the "standard scenario", i.e. inferring tropism from the bulk-sequenced

V3 loop. There we first showed how geno2pheno[coreceptor] can be combined with the WebPSSM-system. Constructing a linear decision boundary from the raw prediction scores of these two freely available web systems yielded a classifier with improved prediction quality. We could demonstrate this improvement in a blinded analysis and subsequently on the large well-defined MOTIVATE-screening cohort. In the second part of this chapter we analyzed prediction results of geno2pheno[coreceptor] in comparison with the enhanced sensitivity Trofile assay which is supposed to have a better sensitivity of X4-detection. We observed decreased sensitivity of geno2pheno[coreceptor] in conjunction with increased specificities. Thus, we were able to achieve similar performance values with predictions achieved on samples screened with the original Trofile assay by adapting the FPR-cutoffs. We then shifted our focus towards different subtypes and introduced a method to predict the subtype of the V3 loop. This proved to work competitively with other methods using the V3 loop as input and was therefore implemented into the geno2pheno[coreceptor] web system. By grouping samples from daily routine into subtype B and non-B isolates we could demonstrate that coreceptor usage prediction performs better for non-B than on subtype B viruses. In the subsequent section we then compared different datasets and found that clinically derived isolates were more similar to each other than to samples of the clonal dataset. Because clonal data is usually used for training of prediction methods, this might pose a problem. Last but not least, we analyzed the performance of geno2pheno[coreceptor] in the clinical setting measuring virological outcome rather than phenotype. We directly compared the results with the ones of the Trofile assay but did not find a clear difference between these. Instead, phenotypic and genotypic results performed equally well suggesting that they both can be used when screening for coreceptor antagonists.

Among others, the results of this thesis have helped to bring genotypic prediction of viral tropism into clinical routine. Nowadays this type of test is considered a reliable tool when monitoring HIV-1 coreceptor usage. In fact, geno2pheno[coreceptor] is now the recommended tool for tropism determination in the German-Austrian treatment guidelines (DAIG, 2009) and has recently also been recommended by the European guidelines group on clinical management of HIV-1 tropism testing (Vandekerckhove et al., 2010). The webserver is heavily used with more than 240,000 predictions performed in the last three years. Recently, we have re-designed the website, incorporated the clinical models described in (Sing et al., 2007) and added the subtype predictor. Furthermore, geno2pheno[454] has been added to the geno2pheno web suite (http://www.geno2pheno.org) and is currently the only one of its kind.

However, one should not forget that the prediction of HIV-1 coreceptor usage is just one piece in this puzzle of searching for the best anti-HIV therapy. As noted in Section 5.5.2, the success of therapy was mostly independent of tropism prediction when the backbone therapy was highly active. On the other hand, in case of a weak backbone, the correct prediction of the phenotype was especially important for therapy outcome. Therefore one should not overemphasize the correct prediction of viral tropism but also head for appropriate selection of the other drugs in the regimen. Only if all pieces - the drugs themselves, their careful selection, and the adherence of the patients - are adequate and compatible, one can expect successful treatment outcomes in a patient.

## Outlook

In the future, one of the main goals will be to combine the individual approaches described in this work. The inclusion of sequence information beyond V3 proved to be effective in improving HIV-1 coreceptor usage prediction. So did the use of massively parallel sequencing of the V3 loop. Combining these two should not be too problematic from a technical point of view but would be quite resource-intensive and would rely on large datasets for training the models. Sufficient data should also ease finding better and more specific features that will allow to analyze treatment outcome more carefully. Furthermore, the incorporation of structural properties and clinical host markers like CD4-cell counts that have been previously shown to be predictive for viral phenotype should be taken into account. But also other markers such as HLA-types might help improving the prediction system.

Properties of the gp41-protein were found to be significantly associated with tropism. The high predictivity of these insertions and mutations certainly warrants further investigation. In particular, their mode of action and their inability to improve predictions when combined with sequence information of the V3 loop seem to be highly interesting topics.

We have seen that some people respond to maraviroc therapy despite having no active background regimen and harboring CXCR4-using viruses. On the one hand this might be the result of some residual activity of the drugs in the backbone but it might also indicate that there are viruses capable of using CXCR4 but only very ineffectively so. In general, this would fit to our understanding of binding to coreceptors. As far as we are concerned, coreceptor usage of a virus is not a binary decision (*can use CXCR4* or *cannot use CXCR4*) but a more finely granulated quantitative value (e.g. *is a very strong binder*, *binds very well*, *binds from time to time*, ..., *cannot bind at all*). Thus, there might be viruses that can use CXCR4 but cannot do so effectively enough especially if there also other reasons for impaired fitness of these strains (e.g. a low replicative capacity). A better understanding might therefore help patients currently not administered with coreceptor antagonists because of a predicted or phenotyped CXCR4-using virus.

Another point which will come into focus with the advent of massively parallel sequencing will be the haplotype structure of the viral quasispecies. When focusing on a relatively small region like in this work we can more or less simply count the frequencies of different (resistance) mutations. However, when analyzing larger regions, a single amplicon will not cover the entire sequence region of interest so that we have to construct the haplotypes from smaller sequence fragments. There are methods available for this task but so far it is not clear to which extent they can also recover the haplotypes of minor variants and how resistance mutations on different amplicons should be dealt with. A related question will be how to handle viral recombination: If we have two variants with different resistance patterns, do these strains recombine so that we need to compute the worst case scenario (a virus with all the resistance mutations of the two original strains) or could we compute the resistance factor for each strain individually as we did in this work?

In this work we have not addressed the ability to infer coreceptor usage from genotype in patients currently being under a successful regimen but who want to switch therapy due to side effects. This is a point at which genotypic approaches should clearly outcompete phenotypic assays as the latter assays usually cannot be performed in these patients because

they have undetectable plasma viral loads. However, proviral DNA can still be derived and sequenced from peripheral blood mononuclear cells (PBMCs). Interpretation works the same as when using viral RNA but one should be a bit more careful because it is (currently) not possible to decide which integrated virus is viable and infective and which is not. First encouraging results have been presented by some groups but more studies will be certainly needed.

The GENO2PHENO[CORECEPTOR] system uses several methods to detect V3 loops and to assess their quality. While the system works well for bulk-sequenced data in general, these methods could and should be improved when dealing with 454-data. We have already changed some parts such as the alignment procedure, however more work will be certainly needed to improve the running time, the space requirement, and the detection and possibly correction of sequencing errors. In addition, because we have applied most of the described approaches for resistance interpretation of the classical drugs targeting the viral polymerase, one should implement a web system for these, too.

# Bibliography


Adamson, C. S. and Freed, E. O. (2010). Novel approaches to inhibiting HIV-1 replication. *Antiviral Res*, 85(1):119–141.

Adamson, C. S., Sakalian, M., Salzwedel, K., and Freed, E. O. (2010). Polymorphisms in Gag spacer peptide 1 confer varying levels of resistance to the HIV- 1maturation inhibitor bevirimat. *Retrovirology*, 7(1):36.

Alkhatib, G. (2009). The biology of CCR5 and CXCR4. *Curr Opin HIV AIDS*, 4(2):96–103.

Alkhatib, G. and Berger, E. A. (2007). HIV coreceptors: from discovery and designation to new paradigms and promise. *Eur J Med Res*, 12(9):375–384.

Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M., and Berger, E. A. (1996). CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science*, 272(5270):1955–1958.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, 215(3):403–410.

Ancelle, R., Bletry, O., Baglin, A. C., Brun-Vezinet, F., Rey, M. A., and Godeau, P. (1987). Long incubation period for HIV-2 infection. *Lancet*, 1(8534):688–689.

Andeweg, A. C., Leeflang, P., Osterhaus, A. D., and Bosch, M. L. (1993). Both the V2 and V3 regions of the human immunodeficiency virus type 1 surface glycoprotein functionally interact with other envelope regions in syncytium formation. *J Virol*, 67(6):3232–3239.

Arhel, N. J. and Kirchhoff, F. (2009). Implications of Nef: host cell interactions in viral persistence and progression to AIDS. *Curr Top Microbiol Immunol*, 339:147–175.

Ariën, K. K., Vanham, G., and Arts, E. J. (2007). Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol*, 5(2):141–151.

Ariyoshi, K., Jaffar, S., Alabi, A. S., Berry, N., van der Loeff, M. S., Sabally, S., N'Gom, P. T., Corrah, T., Tedder, R., and Whittle, H. (2000). Plasma RNA viral load predicts the rate of CD4 T cell decline and death in HIV-2-infected patients in West Africa. *AIDS*, 14(4):339–344.

Arthos, J., Cicala, C., Martinelli, E., Macleod, K., Ryk, D. V., Wei, D., Xiao, Z., Veenstra, T. D., Conrad, T. P., Lempicki, R. A., McLaughlin, S., Pascuccio, M., Gopaul, R., McNally, J., Cruz, C. C., Censoplano, N., Chung, E., Reitano, K. N., Kottilil, S., Goode, D. J., and Fauci, A. S. (2008). HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells. *Nat Immunol*, 9(3):301–309.

Ashorn, P. A., Berger, E. A., and Moss, B. (1990). Human immunodeficiency virus envelope glycoprotein/CD4-mediated fusion of nonprimate cells with human cells. *J Virol*, 64(5):2149–2156.

Asjö, B., Morfeldt-Månson, L., Albert, J., Biberfeld, G., Karlsson, A., Lidman, K., and Fenyö, E. M. (1986). Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. *Lancet*, 2(8508):660–662.

Back, N. K., Smit, L., Schutten, M., Nara, P. L., Tersmette, M., and Goudsmit, J. (1993). Mutations in human immunodeficiency virus type 1 gp41 affect sensitivity to neutralization by gp120 antibodies. *J Virol*, 67(11):6897–6902.

Baelen, K. V., Vandenbroucke, I., Rondelez, E., Eygen, V. V., Vermeiren, H., and Stuyver, L. J. (2007). HIV-1 coreceptor usage determination in clinical isolates using clonal and population-based genotypic and phenotypic assays. *J Virol Methods*, 146(1-2):61–73.

Baeten, J. M., Chohan, B., Lavreys, L., Chohan, V., McClelland, R. S., Certain, L., Mandaliya, K., Jaoko, W., and Overbaugh, J. (2007). HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J Infect Dis*, 195(8):1177–1180.

Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, 220(4599):868–871.

Beerenwinkel, N., Nolden, T., Kupfer, B., Selbig, J., Daeumer, M., Hoffmann, D., R., J. R., and Kaiser (2001). Predicting HIV-1 cytopathogenicity and co-receptor usage from V3 envelope sequences by machine learning. 1st IAS Conf HIV Pathog Treat, Abstract No. 268.

Bell, C., Devarajan, S., and Gersbach, H. (2003). The long-run economic costs of AIDS: theory and application to South Africa. Technical report, World Bank, Washington, DC.

Berger, E. A., Doms, R. W., Fenyö, E. M., Korber, B. T., Littman, D. R., Moore, J. P., Sattentau, Q. J., Schuitemaker, H., Sodroski, J., and Weiss, R. A. (1998). A new classification for HIV-1. *Nature*, 391(6664):240.

Berger, E. A., Murphy, P. M., and Farber, J. M. (1999). Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol*, 17:657–700.

Berry, N., Ariyoshi, K., Jaffar, S., Sabally, S., Corrah, T., Tedder, R., and Whittle, H. (1998). Low peripheral blood viral HIV-2 RNA in individuals with high CD4 percentage differentiates HIV-2 from HIV-1 infection. *J Hum Virol*, 1(7):457–468.

Bhattacharya, S. and Osman, H. (2009). Novel targets for anti-retroviral therapy. *J Infect*, 59(6):377–386.

Bickel, S., Bogojeska, J., Lengauer, T., and Scheffer, T. (2008). Multi-task learning for HIV therapy screening. In *ICML '08: Proceedings of the 25th international conference on Machine learning*, pages 56–63, New York, NY, USA. ACM.

Biscone, M. J., Miamidian, J. L., Muchiri, J. M., Baik, S. S. W., Lee, F.-H., Doms, R. W., and Reeves, J. D. (2006). Functional impact of HIV coreceptor-binding site mutations. *Virology*, 351(1):226–236.

Blish, C. A., Nguyen, M.-A., and Overbaugh, J. (2008). Enhancing exposure of HIV-1 neutralization epitopes through mutations in gp41. *PLoS Med*, 5(1):e9.

Boisvert, S., Marchand, M., Laviolette, F., and Corbeil, J. (2008). HIV-1 coreceptor usage prediction without multiple alignments: an application of string kernels. *Retrovirology*, 5(1):110.

Borrow, P., Lewicki, H., Hahn, B. H., Shaw, G. M., and Oldstone, M. B. (1994). Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *J Virol*, 68(9):6103–6110.

Boyd, M. T., Simpson, G. R., Cann, A. J., Johnson, M. A., and Weiss, R. A. (1993). A single amino acid substitution in the V1 loop of human immunodeficiency virus type 1 gp120 alters cellular tropism. *J Virol*, 67(6):3649–3652.

Brennan, R. O. and Durack, D. T. (1981). Gay compromise syndrome. *Lancet*, 2(8259):1338–1339.

Briggs, D. R., Tuttle, D. L., Sleasman, J. W., and Goodenow, M. M. (2000). Envelope V3 amino acid sequence predicts HIV-1 phenotype (co-receptor usage and tropism for macrophages). *AIDS*, 14(18):2937–2939.

Brooks, J. I., Woods, C. K., Merks, H., Wynhoven, B., Hall, T. A., Sandstrom, P. A., and Harrigan, P. R. (2009). Evaluation of an Automated Sequence Analysis Tool to Standardize HIV Genotyping Results. In *18th Annual Canadian Conference on HIV/AIDS Research, April 23-26, 2009. Vancouver, Canada.*

Brumme, Z. L., Goodrich, J., Mayer, H. B., Brumme, C. J., Henrick, B. M., Wynhoven, B., Asselin, J. J., Cheung, P. K., Hogg, R. S., Montaner, J. S. G., and Harrigan, P. R. (2005). Molecular and clinical epidemiology of CXCR4-using HIV-1 in a large population of antiretroviral-naive individuals. *J Infect Dis*, 192(3):466–474.

Buchbinder, S. P., Mehrotra, D. V., Duerr, A., Fitzgerald, D. W., Mogg, R., Li, D., Gilbert, P. B., Lama, J. R., Marmor, M., Rio, C. D., McElrath, M. J., Casimiro, D. R., Gottesdiener, K. M., Chodakewitz, J. A., Corey, L., Robertson, M. N., and Team, S. S. P. (2008). Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *Lancet*, 372(9653):1881–1893.

Butler, I. F., Pandrea, I., Marx, P. A., and Apetrei, C. (2007). HIV genetic diversity: biological and public health consequences. *Curr HIV Res*, 5(1):23–45.

Carrillo, A. and Ratner, L. (1996). Human immunodeficiency virus type 1 tropism for T-lymphoid cell lines: role of the V3 loop and C4 envelope determinants. *J Virol*, 70(2):1301–1309.

Chang, C.-C. and Lin, C.-J. (2001). *LIBSVM: a library for support vector machines*. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P. D., Wu, L., Mackay, C. R., LaRosa, G., Newman, W., Gerard, N., Gerard, C., and Sodroski, J. (1996). The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell*, 85(7):1135–1148.

Chohan, B., Lang, D., Sagar, M., Korber, B., Lavreys, L., Richardson, B., and Overbaugh, J. (2005). Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. *J Virol*, 79(10):6528–6531.

Choo, V. (1995). Combination superior to zidovudine in Delta trial. *Lancet*, 346(8979):895.

Chueca, N., Garrido, C., Alvarez, M., Poveda, E., de Dios Luna, J., Zahonero, N., Hernández-Quero, J., Soriano, V., Maroto, C., de Mendoza, C., and García, F. (2009). Improvement in the determination of HIV-1 tropism using the V3 gene sequence and a combination of bioinformatic tools. *J Med Virol*, 81(5):763–767.

Chun, T.-W., Nickle, D. C., Justement, J. S., Meyers, J. H., Roby, G., Hallahan, C. W., Kottilil, S., Moir, S., Mican, J. M., Mullins, J. I., Ward, D. J., Kovacs, J. A., Mannon, P. J., and Fauci, A. S. (2008). Persistence of HIV in gut-associated lymphoid tissue despite long-term antiretroviral therapy. *J Infect Dis*, 197(5):714–720.

Cihlar, T. and Ray, A. S. (2010). Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Res*, 85(1):39–58.

Clapham, P. R., Blanc, D., and Weiss, R. A. (1991). Specific cell surface requirements for the infection of CD4-positive cells by human immunodeficiency virus types 1 and 2 and by Simian immunodeficiency virus. *Virology*, 181(2):703–715.

Clavel, F. and Hance, A. J. (2004). HIV drug resistance. *N Engl J Med*, 350(10):1023–1035.

Clavel, F., Mansinho, K., Chamaret, S., Guetard, D., Favier, V., Nina, J., Santos-Ferreira, M. O., Champalimaud, J. L., and Montagnier, L. (1987). Human immunodeficiency virus type 2 infection associated with AIDS in West Africa. *N Engl J Med*, 316(19):1180–1185.

Coakley, E., Petropoulos, C. J., and Whitcomb, J. M. (2005). Assessing chemokine co-receptor usage in HIV. *Curr Opin Infect Dis*, 18(1):9–15.

Coakley, E., Reeves, J. D., Huang, W., Mangas-Ruiz, M., Maurer, I., Harskamp, A. M., Gupta, S., Lie, Y., Petropoulos, C. J., Schuitemaker, H., and van 't Wout, A. B. (2009). Comparison of human immunodeficiency virus type 1 tropism profiles in clinical samples by the Trofile and MT-2 assays. *Antimicrob Agents Chemother*, 53(11):4686–4693.

Cocchi, F., DeVico, A. L., Garzino-Demo, A., Arya, S. K., Gallo, R. C., and Lusso, P. (1995). Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science*, 270(5243):1811–1815.

Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., and Vogt, P. (1986). What to call the AIDS virus? *Nature*, 321(6065):10.

Combadiere, C., Ahuja, S. K., Tiffany, H. L., and Murphy, P. M. (1996). Cloning and functional expression of CC CKR5, a human monocyte CC chemokine receptor selective for MIP-1(alpha), MIP-1(beta), and RANTES. *J Leukoc Biol*, 60(1):147–152.

Connor, R. I. and Ho, D. D. (1994). Transmission and pathogenesis of human immunodeficiency virus type 1. *AIDS Res Hum Retroviruses*, 10(4):321–323.

Cooper, D. A., Heera, J., Goodrich, J., Tawadrous, M., Saag, M., Dejesus, E., Clumeck, N., Walmsley, S., Ting, N., Coakley, E., Reeves, J. D., Reyes-Teran, G., Westby, M., Ryst, E. V. D., Ive, P., Mohapi, L., Mingrone, H., Horban, A., Hackman, F., Sullivan, J., and Mayer, H. (2010). Maraviroc versus Efavirenz, Both in Combination with Zidovudine-Lamivudine, for the Treatment of Antiretroviral-Naive Subjects with CCR5-Tropic HIV-1 Infection. *J Infect Dis*.

Cullen, B. R. (1991). Regulation of HIV-1 gene expression. *FASEB J*, 5(10):2361–2368.

Daar, E. S., Kesler, K. L., Petropoulos, C. J., Huang, W., Bates, M., Lail, A. E., Coakley, E. P., Gomperts, E. D., Donfield, S. M., Growth, H., and Study, D. (2007). Baseline HIV type 1 coreceptor tropism predicts disease progression. *Clin Infect Dis*, 45(5):643–649.

DAIG (2009). Leitlinien zur Topismus-Testung. Technical report, Deutsche und Österreichische AIDS-Gesellschaft.

Dalgleish, A. G., Beverley, P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F., and Weiss, R. A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*, 312(5996):763–767.

Däumer, M., Kaiser, R., Klein, R., Lengauer, T., Thiele, B., and Thielen, A. (2008). Inferring viral tropism from genotype with massively parallel sequencing; qualitative and quantitative analysis. In *Antiviral Therapy*, volume 13, page A101, Sitges, Spain. International Medical, international medical.

de Béthune, M.-P. (2010). Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989-2009). *Antiviral Res*, 85(1):75–90.

de Mendoza, C., Baelen, K. V., Poveda, E., Rondelez, E., Zahonero, N., Stuyver, L., Garrido, C., Villacian, J., Soriano, V., and Group, S. H. S. S. (2008). 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*, 48(3):241–244.

de Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., Snoeck, J., van Rensburg, E. J., Wensing, A. M. J., van de Vijver, D. A., Boucher, C. A., Camacho, R., and Vandamme, A.-M. (2005). An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics*, 21(19):3797–3800.

de Silva, T. I., Cotten, M., and Rowland-Jones, S. L. (2008). HIV-2: the forgotten AIDS virus. *Trends Microbiol*, 16(12):588–595.

Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., Goedert, J. J., Buchbinder, S. P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R., and O'Brien, S. J. (1996). Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science*, 273(5283):1856–1862.

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3):837–845.

Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Marzio, P. D., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R., and Landau, N. R. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, 381(6584):661–666.

Derdeyn, C. A., Decker, J. M., Bibollet-Ruche, F., Mokili, J. L., Muldoon, M., Denham, S. A., Heil, M. L., Kasolo, F., Musonda, R., Hahn, B. H., Shaw, G. M., Korber, B. T., Allen, S., and Hunter, E. (2004). Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science*, 303(5666):2019–2022.

Derdeyn, C. A., Decker, J. M., Sfakianos, J. N., Wu, X., O'Brien, W. A., Ratner, L., Kappes, J. C., Shaw, G. M., and Hunter, E. (2000). Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. *J Virol*, 74(18):8358–8367.

Derdeyn, C. A., Decker, J. M., Sfakianos, J. N., Zhang, Z., O'Brien, W. A., Ratner, L., Shaw, G. M., and Hunter, E. (2001). Sensitivity of human immunodeficiency virus type 1 to fusion inhibitors targeted to the gp41 first heptad repeat involves distinct regions of gp41 and is consistently modulated by gp120 interactions with the coreceptor. *J Virol*, 75(18):8605–8614.

Descamps, D., Apetrei, C., Collin, G., Damond, F., Simon, F., and Brun-Vézinet, F. (1998). Naturally occurring decreased susceptibility of HIV-1 subtype G to protease inhibitors. *AIDS*, 12(9):1109–1111.

Doranz, B. J., Rucker, J., Yi, Y., Smyth, R. J., Samson, M., Peiper, S. C., Parmentier, M., Collman, R. G., and Doms, R. W. (1996). 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*, 85(7):1149–1158.

Dorr, P., Westby, M., Dobbs, S., Griffin, P., Irvine, B., Macartney, M., Mori, J., Rickett, G., Smith-Burchnell, C., Napier, C., Webster, R., Armour, D., Price, D., Stammen, B., Wood, A., and Perros, M. (2005). 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*, 49(11):4721–4732.

Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., Cayanan, C., Maddon, P. J., Koup, R. A., Moore, J. P., and Paxton, W. A. (1996). HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature*, 381(6584):667–673.

D'Souza, V. and Summers, M. F. (2005). How retroviruses select their genomes. *Nat Rev Microbiol*, 3(8):643–655.

Dumond, J. B., Patterson, K. B., Pecha, A. L., Werner, R. E., Andrews, E., Damle, B., Tressler, R., Worsley, J., and Kashuba, A. D. M. (2009). Maraviroc concentrates in the cervicovaginal fluid and vaginal tissue of HIV-negative women. *J Acquir Immune Defic Syndr*, 51(5):546–553.

Dybowski, J. N., Heider, D., and Hoffmann, D. (2010). Prediction of co-receptor usage of HIV-1 from genotype. *PLoS Comput Biol*, 6(4):e1000743.

Esté, J. A. and Cihlar, T. (2010). Current status and challenges of antiretroviral research and therapy. *Antiviral Res*, 85(1):25–33.

Esté, J. A. and Telenti, A. (2007). HIV entry inhibitors. *Lancet*, 370(9581):81–88.

Fauci, A. S. (1996). Host factors and the pathogenesis of HIV-induced disease. *Nature*, 384(6609):529–534.

Fellay, J., Shianna, K. V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., Cozzi-Lepri, A., Luca, A. D., Easterbrook, P., Francioli, P., Mallal, S., Martinez-Picado, J., Miro, J. M., Obel, N., Smith, J. P., Wyniger, J., Descombes, P., Antonarakis, S. E., Letvin, N. L., McMichael, A. J., Haynes, B. F., Telenti, A., and Goldstein, D. B. (2007). A whole-genome association study of major determinants for host control of HIV-1. *Science*, 317(5840):944–947.

Feng, Y., Broder, C. C., Kennedy, P. E., and Berger, E. A. (1996). HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science*, 272(5263):872–877.

Fenyö, E. M., Fiore, J., Karlsson, A., Albert, J., and Scarlatti, G. (1994). Biological phenotypes of HIV-1 in pathogenesis and transmission. *Antibiot Chemother*, 46:18–24.

Fenyö, E. M., Morfeldt-Månson, L., Chiodi, F., Lind, B., von Gegerfelt, A., Albert, J., Olausson, E., and Asjö, B. (1988). Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. *J Virol*, 62(11):4414–4419.

Fischl, M. A., Richman, D. D., Grieco, M. H., Gottlieb, M. S., Volberding, P. A., Laskin, O. L., Leedom, J. M., Groopman, J. E., Mildvan, D., and Schooley, R. T. (1987). The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial. *N Engl J Med*, 317(4):185–191.

for Disease Control, C. (1981). Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men–New York City and California. *MMWR Morb Mortal Wkly Rep*, 30(25):305–308.

Fouchier, R. A., Groenink, M., Kootstra, N. A., Tersmette, M., Huisman, H. G., Miedema, F., and Schuitemaker, H. (1992). Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol*, 66(5):3183–3187.

Frankel, A. D. and Young, J. A. (1998). HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem*, 67:1–25.

Freed, E. O. (2001). HIV-1 replication. *Somat Cell Mol Genet*, 26(1-6):13–33.

Frey, G., Peng, H., Rits-Volloch, S., Morelli, M., Cheng, Y., and Chen, B. (2008). A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *Proc Natl Acad Sci U S A*, 105(10):3739–3744.

Fätkenheuer, G., Nelson, M., Lazzarin, A., Konourina, I., Hoepelman, A. I. M., Lampiris, H., Hirschel, B., Tebas, P., Raffi, F., Trottier, B., Bellos, N., Saag, M., Cooper, D. A., Westby, M., Tawadrous, M., Sullivan, J. F., Ridgway, C., Dunne, M. W., Felstead, S., Mayer, H., van der Ryst, E., 1, M. O. T. I. V. A. T. E., and Teams, M. O. T. I. V. A. T. E. . S. (2008). Subgroup analyses of maraviroc in previously treated R5 HIV-1 infection. *N Engl J Med*, 359(14):1442–1455.

Gallo, R. C. and Montagnier, L. (2003). The discovery of HIV as the cause of AIDS. *N Engl J Med*, 349(24):2283–2285.

Garrido, C., Roulet, V., Chueca, N., Poveda, E., Aguilera, A., Skrabal, K., Zahonero, N., Carlos, S., García, F., Faudon, J. L., Soriano, V., and de Mendoza, C. (2008a). Evaluation of eight different bioinformatics tools to predict viral tropism in different human immunodeficiency virus type 1 subtypes. *J Clin Microbiol*, 46(3):887–891.

Garrido, C., Roulet, V., Chueca, N., Poveda, E., Aguilera, A., Skrabal, K., Zahonero, N., Carlos, S., García, F., Faudon, J. L., Soriano, V., and de Mendoza, C. (2008b). Evaluation of Eight Different Bioinformatics Tools to Predict Viral Tropism in Different HIV-1 Subtypes. *J Clin Microbiol*.

Gelderblom, H. R. (1991). Assembly and morphology of HIV: potential effect of structure on viral function. *AIDS*, 5(6):617–637.

Gottlieb, G. S., Sow, P. S., Hawes, S. E., Ndoye, I., Redman, M., Coll-Seck, A. M., Faye-Niang, M. A., Diop, A., Kuypers, J. M., Critchlow, C. W., Respess, R., Mullins, J. I., and Kiviat, N. B. (2002). Equal plasma viral loads predict a similar rate of CD4+ T cell decline in human immunodeficiency virus (HIV) type 1- and HIV-2-infected individuals from Senegal, West Africa. *J Infect Dis*, 185(7):905–914.

Gottlieb, M. S. (2006). Pneumocystis pneumonia–Los Angeles. 1981. *Am J Public Health*, 96(6):980–1; discussion 982–3.

Grivel, J. C. and Margolis, L. B. (1999). CCR5- and CXCR4-tropic HIV-1 are equally cytopathic for their T-cell targets in human lymphoid tissue. *Nat Med*, 5(3):344–346.

Groenink, M., Fouchier, R. A., Broersen, S., Baker, C. H., Koot, M., van't Wout, A. B., Huisman, H. G., Miedema, F., Tersmette, M., and Schuitemaker, H. (1993). Relation of phenotype evolution of HIV-1 to envelope V2 configuration. *Science*, 260(5113):1513–1516.

Gulick, R. M., Lalezari, J., Goodrich, J., Clumeck, N., DeJesus, E., Horban, A., Nadler, J., Clotet, B., Karlsson, A., Wohlfeiler, M., Montana, J. B., McHale, M., Sullivan, J., Ridgway, C., Felstead, S., Dunne, M. W., van der Ryst, E., Mayer, H., and Teams, M. O. T. I. V. A. T. E. S. (2008). Maraviroc for previously treated patients with R5 HIV-1 infection. *N Engl J Med*, 359(14):1429–1441.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S., and Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. AIDS Clinical Trials Group Study 175 Study Team. *N Engl J Med*, 335(15):1081–1090.

Hansmann, A., van der Loeff, M. F. S., Kaye, S., Awasana, A. A., Sarge-Njie, R., O'Donovan, D., Ariyoshi, K., Alabi, A., Milligan, P., and Whittle, H. C. (2005). Baseline plasma viral load and CD4 cell percentage predict survival in HIV-1- and HIV-2-infected women in a community-based cohort in The Gambia. *J Acquir Immune Defic Syndr*, 38(3):335–341.

Harrigan, P. R., Hogg, R. S., Dong, W. W. Y., Yip, B., Wynhoven, B., Woodward, J., Brumme, C. J., Brumme, Z. L., Mo, T., Alexander, C. S., and Montaner, J. S. G. (2005). Predictors of HIV drug-resistance mutations in a large antiretroviral-naive cohort initiating triple antiretroviral therapy. *J Infect Dis*, 191(3):339–347.

Hawkins, T. (2010). Understanding and managing the adverse effects of antiretroviral therapy. *Antiviral Res*, 85(1):201–209.

Hirsch, M. S. (1990). Chemotherapy of human immunodeficiency virus infections: current practice and future prospects. *J Infect Dis*, 161(5):845–857.

Hoffman, N. G., Seillier-Moiseiwitsch, F., Ahn, J., Walker, J. M., and Swanstrom, R. (2002). Variability in the human immunodeficiency virus type 1 gp120 Env protein linked to phenotype-associated changes in the V3 loop. *J Virol*, 76(8):3852–3864.

Hogg, R. S., Heath, K. V., Yip, B., Craib, K. J., O'Shaughnessy, M. V., Schechter, M. T., and Montaner, J. S. (1998a). Improved survival among HIV-infected individuals following initiation of antiretroviral therapy. *JAMA*, 279(6):450–454.

Hogg, R. S., Rhone, S. A., Yip, B., Sherlock, C., Conway, B., Schechter, M. T., O'Shaughnessy, M. V., and Montaner, J. S. (1998b). Antiviral effect of double and triple drug combinations amongst HIV-infected adults: lessons from the implementation of viral load-driven antiretroviral therapy. *AIDS*, 12(3):279–284.

Hogg, R. S., Yip, B., Chan, K. J., Wood, E., Craib, K. J., O'Shaughnessy, M. V., and Montaner, J. S. (2001). Rates of disease progression by baseline CD4 cell count and viral load after initiating triple-drug therapy. *JAMA*, 286(20):2568–2577.

Horsburgh, C. R. and Holmberg, S. D. (1988). The global distribution of human immunodeficiency virus type 2 (HIV-2) infection. *Transfusion*, 28(2):192–195.

Hütter, G., Nowak, D., Mossner, M., Ganepola, S., Müssig, A., Allers, K., Schneider, T., Hofmann, J., Kücherer, C., Blau, O., Blau, I. W., Hofmann, W. K., and Thiel, E. (2009). Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med*, 360(7):692–698.

Huang, W., Eshleman, S. H., Toma, J., Fransen, S., Stawiski, E., Paxinos, E. E., Whitcomb, J. M., Young, A. M., Donnell, D., Mmiro, F., Musoke, P., Guay, L. A., Jackson, J. B., Parkin, N. T., and Petropoulos, C. J. (2007). Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous composition of viral populations. *J Virol*, 81(15):7885–7893.

Huang, W., Toma, J., Fransen, S., Stawiski, E., Reeves, J. D., Whitcomb, J. M., Parkin, N., and Petropoulos, C. J. (2008). Coreceptor Tropism can be Influenced by Amino Acid Substitutions in the gp41 Transmembrane Subunit of Human Immunodeficiency Virus Type 1 Envelope Protein. *J Virol*.

Hwang, S. S., Boyle, T. J., Lyerly, H. K., and Cullen, B. R. (1991). Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science*, 253(5015):71–74.

Hymes, K. B., Cheung, T., Greene, J. B., Prose, N. S., Marcus, A., Ballard, H., William, D. C., and Laubenstein, L. J. (1981). Kaposi's sarcoma in homosexual men-a report of eight cases. *Lancet*, 2(8247):598–600.

Invernizzi, C. F., Coutsinos, D., Oliveira, M., Moisi, D., Brenner, B. G., and Wainberg, M. A. (2009). Signature nucleotide polymorphisms at positions 64 and 65 in reverse transcriptase favor the selection of the K65R resistance mutation in HIV-1 subtype C. *J Infect Dis*, 200(8):1202–1206.

Jensen, M. A., Coetzer, M., van 't Wout, A. B., Morris, L., and Mullins, J. I. (2006). A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences. *J Virol*, 80(10):4698–4704.

Jensen, M. A., Li, F.-S., van 't Wout, A. B., Nickle, D. C., Shriner, D., He, H.-X., McLaughlin, S., Shankarappa, R., Margolick, J. B., and Mullins, J. I. (2003). Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *J Virol*, 77(24):13376–13388.

Joachims, T. (1998). Making large-scale support vector machine learning practical. In Schölkopf, B., Burges, C., and Smola, A., editors, *Advances in Kernel Methods: Support Vector Machines*. MIT Press, Cambridge, MA.

Johnson, V. A., Brun-Vezinet, F., Clotet, B., Gunthard, H. F., Kuritzkes, D. R., Pillay, D., Schapiro, J. M., and Richman, D. D. (2008). Update of the Drug Resistance Mutations in HIV-1. *Top HIV Med*, 16(5):138–145.

Jung, A., Maier, R., Vartanian, J.-P., Bocharov, G., Jung, V., Fischer, U., Meese, E., Wain-Hobson, S., and Meyerhans, A. (2002). Recombination: Multiply infected spleen cells in HIV patients. *Nature*, 418(6894):144.

Kalia, V., Sarkar, S., Gupta, P., and Montelaro, R. C. (2003). Rational site-directed mutations of the LLP-1 and LLP-2 lentivirus lytic peptide domains in the intracytoplasmic tail of human immunodeficiency virus type 1 gp41 indicate common functions in cell-cell fusion but distinct roles in virion envelope incorporation. *J Virol*, 77(6):3634–3646.

Kim, J. B. and Sharp, P. A. (2001). Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. *J Biol Chem*, 276(15):12317–12323.

Kiwanuka, N., Robb, M., Laeyendecker, O., Kigozi, G., Wabwire-Mangen, F., Makumbi, F. E., Nalugoda, F., Kagaayi, J., Eller, M., Eller, L. A., Serwadda, D., Sewankambo, N. K., Reynolds, S. J., Quinn, T. C., Gray, R. H., Wawer, M. J., and Whalen, C. C. (2009). HIV-1 Viral Subtype Differences in the Rate of CD4+ T-Cell Decline Among HIV Seroincident Antiretroviral Naive Persons in Rakai District, Uganda. *J Acquir Immune Defic Syndr*.

Koito, A., Harrowe, G., Levy, J. A., and Cheng-Mayer, C. (1994). Functional role of the V1/V2 region of human immunodeficiency virus type 1 envelope glycoprotein gp120 in infection of primary macrophages and soluble CD4 neutralization. *J Virol*, 68(4):2253–2259.

Koito, A., Stamatatos, L., and Cheng-Mayer, C. (1995). Small amino acid sequence changes within the V2 domain can affect the function of a T-cell line-tropic human immunodeficiency virus type 1 envelope gp120. *Virology*, 206(2):878–884.

Koot, M., Keet, I. P., Vos, A. H., de Goede, R. E., Roos, M. T., Coutinho, R. A., Miedema, F., Schellekens, P. T., and Tersmette, M. (1993). Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. *Ann Intern Med*, 118(9):681–688.

Koot, M., Vos, A. H., Keet, R. P., de Goede, R. E., Dercksen, M. W., Terpstra, F. G., Coutinho, R. A., Miedema, F., and Tersmette, M. (1992). HIV-1 biological phenotype in long-term infected individuals evaluated with an MT-2 cocultivation assay. *AIDS*, 6(1):49–54.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S., and Bhattacharya, T. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science*, 288(5472):1789–1796.

Kreisberg, J. F., Kwa, D., Schramm, B., Trautner, V., Connor, R., Schuitemaker, H., Mullins, J. I., van't Wout, A. B., and Goldsmith, M. A. (2001). Cytopathicity of human

immunodeficiency virus type 1 primary isolates depends on coreceptor usage and not patient disease status. *J Virol*, 75(18):8842–8847.

Kupfer, B., Kaiser, R., Rockstroh, J. K., Matz, B., and Schneweis, K. E. (1998). Role of HIV-1 phenotype in viral pathogenesis and its relation to viral load and CD4+ T-cell count. *J Med Virol*, 56(3):259–263.

Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, 393(6686):648–659.

Labrosse, B., Labernardière, J.-L., Dam, E., Trouplin, V., Skrabal, K., Clavel, F., and Mammano, F. (2003). Baseline susceptibility of primary human immunodeficiency virus type 1 to entry inhibitors. *J Virol*, 77(2):1610–1613.

Larder, B. A. and Kemp, S. D. (1989). Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science*, 246(4934):1155–1158.

Larder, B. A., Kemp, S. D., and Purifoy, D. J. (1989). Infectious potential of human immunodeficiency virus type 1 reverse transcriptase mutants with altered inhibitor sensitivity. *Proc Natl Acad Sci U S A*, 86(13):4803–4807.

Lengauer, T., Sander, O., Sierra, S., Thielen, A., and Kaiser, R. (2007). Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol*, 25(12):1407–1410.

Lieberman-Blum, S. S., Fung, H. B., and Bandres, J. C. (2008). Maraviroc: a CCR5-receptor antagonist for the treatment of HIV-1 infection. *Clin Ther*, 30(7):1228–1250.

Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R., MacDonald, M. E., Stuhlmann, H., Koup, R. A., and Landau, N. R. (1996). Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell*, 86(3):367–377.

Liu, Y., Curlin, M. E., Diem, K., Zhao, H., Ghosh, A. K., Zhu, H., Woodward, A. S., Maenza, J., Stevens, C. E., Stekler, J., Collier, A. C., Genowati, I., Deng, W., Zioni, R., Corey, L., Zhu, T., and Mullins, J. I. (2008). Env length and N-linked glycosylation following transmission of human immunodeficiency virus Type 1 subtype B viruses. *Virology*, 374(2):229–233.

Livingstone, C. D. and Barton, G. J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Computer applications in the biosciences : CABIOS*, 9(6):745–756.

Low, A., Dong, W., Chan, D., Sing, T., Swanstrom, R., Jensen, M., Pillai, S., Good, B., and Harrigan, P. (2007). Current V3 genotyping algorithms are inadequate for predicting X4 co-receptor usage in clinical isolates. *AIDS*, 21(14):F17–F24.

Low, A. J., Swenson, L. C., and Harrigan, P. R. (2008). HIV Coreceptor Phenotyping in the Clinical Setting. *AIDS Rev*, 10(3):143–151.

Lyles, R. H., Muñoz, A., Yamashita, T. E., Bazmi, H., Detels, R., Rinaldo, C. R., Margolick, J. B., Phair, J. P., and Mellors, J. W. (2000). Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study. *J Infect Dis*, 181(3):872–880.

Maddon, P. J., Dalgleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A., and Axel, R. (1986). The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell*, 47(3):333–348.

Martínez-Cajas, J. L., Pant-Pai, N., Klein, M. B., and Wainberg, M. A. (2008). Role of genetic diversity amongst HIV-1 non-B subtypes in drug resistance: a systematic review of virologic and biochemical evidence. *AIDS Rev*, 10(4):212–223.

Mattapallil, J. J., Douek, D. C., Hill, B., Nishimura, Y., Martin, M., and Roederer, M. (2005). Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. *Nature*, 434(7037):1093–1097.

McColl, D. J. and Chen, X. (2010). Strand transfer inhibitors of HIV-1 integrase: bringing IN a new era of antiretroviral therapy. *Antiviral Res*, 85(1):101–118.

McGovern, R. A., Thielen, A., Mo, T., Dong, W., Woods, C. K., Chapman, D., Lewis, M., James, I., Heera, J., Valdez, H., and Harrigan, P. R. (2010). Population-based V3 genotypic tropism assay: a retrospective analysis using screening samples from the A4001029 and MOTIVATE studies. *AIDS*, 24(16):2517–2525.

McMichael, A. J. (2006). HIV vaccines. *Annu Rev Immunol*, 24:227–255.

Melby, T., Despirito, M., Demasi, R., Heilek-Snyder, G., Greenberg, M. L., and Graham, N. (2006). HIV-1 coreceptor use in triple-class treatment-experienced patients: baseline prevalence, correlates, and relationship to enfuvirtide response. *J Infect Dis*, 194(2):238–246.

Mellors, J. W., Rinaldo, C. R., Gupta, P., White, R. M., Todd, J. A., and Kingsley, L. A. (1996). Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science*, 272(5265):1167–1170.

Menéndez-Arias, L. (2010). Molecular basis of human immunodeficiency virus drug resistance: an update. *Antiviral Res*, 85(1):210–231.

Miedema, F., Meyaard, L., Koot, M., Klein, M. R., Roos, M. T., Groenink, M., Fouchier, R. A., Wout, A. B. V., Tersmette, M., and Schellekens, P. T. (1994). Changing virus-host interactions in the course of HIV-1 infection. *Immunol Rev*, 140:35–72.

Moore, J. P. and Kuritzkes, D. R. (2009). A pièce de resistance: how HIV-1 escapes small molecule CCR5 inhibitors. *Curr Opin HIV AIDS*, 4(2):118–124.

Moores, A., Thielen, A., Dong, W., Low, A. J., Woods, C., Jensen, M., Wynhoven, B., Chan, D., Glascock, C., and Harrigan, P. R. (2008). Improved detection of X4 virus by V3 genotyping: application to plasma RNA and proviral DNA. In *Antiviral Therapy*, volume 13, page A99, Sitges, Spain. international medical.

Mougel, M., Houzet, L., and Darlix, J.-L. (2009). When is it time for reverse transcription to start and go? *Retrovirology*, 6:24.

Moulard, M. and Decroly, E. (2000). Maturation of HIV envelope glycoprotein precursors by cellular endoproteases. *Biochim Biophys Acta*, 1469(3):121–132.

Murphy, P. M., Baggiolini, M., Charo, I. F., Hébert, C. A., Horuk, R., Matsushima, K., Miller, L. H., Oppenheim, J. J., and Power, C. A. (2000). International union of pharmacology. XXII. Nomenclature for chemokine receptors. *Pharmacol Rev*, 52(1):145–176.

Nedellec, R., Coetzer, M., Shimizu, N., Hoshino, H., Polonis, V. R., Morris, L., Mårtensson, U. E. A., Binley, J., Overbaugh, J., and Mosier, D. E. (2009). Virus entry via the alternative coreceptors CCR3 and FPRL1 differs by human immunodeficiency virus type 1 subtype. *J Virol*, 83(17):8353–8363.

Neil, S. J. D., Sandrin, V., Sundquist, W. I., and Bieniasz, P. D. (2007). An interferon-alpha-induced tethering mechanism inhibits HIV-1 and Ebola virus particle release but is counteracted by the HIV-1 Vpu protein. *Cell Host Microbe*, 2(3):193–203.

Nelson, K. E., Costello, C., Suriyanon, V., Sennun, S., and Duerr, A. (2007). Survival of blood donors and their spouses with HIV-1 subtype E (CRF01 AE) infection in northern Thailand, 1992-2007. *AIDS*, 21 Suppl 6:S47–S54.

Nijhuis, M., van Maarseveen, N. M., Verheyen, J., and Boucher, C. A. (2008). Novel mechanisms of HIV protease inhibitor resistance. *Curr Opin HIV AIDS*, 3(6):627–632.

Novembre, J., Galvani, A. P., and Slatkin, M. (2005). The geographic spread of the CCR5 Delta32 HIV-resistance allele. *PLoS Biol*, 3(11):e339.

Obermeier, M., Berg, T., Sichtig, N., Braun, P., Däumer, M., Walter, H., Noah, C., Wolf, E., Müller, H., Stürmer, M., Thielen, A., and Kaiser, R. (2008). Determination of HIV-1 co-receptor usage in German patients - comparison of genotypic methods with the TROFILEő phenotypic assay. In *Abstracts of the Ninth International Congress on Drug Therapy in HIV Infection*, volume 11 Suppl 1, page 123, Glasgow, UK. BioMed Central.

O'Donovan, D., Ariyoshi, K., Milligan, P., Ota, M., Yamuah, L., Sarge-Njie, R., and Whittle, H. (2000). Maternal plasma viral RNA levels determine marked differences in mother-to-child transmission rates of HIV-1 and HIV-2 in The Gambia. MRC/Gambia Government/University College London Medical School working group on mother-child transmission of HIV. *AIDS*, 14(4):441–448.

Pancera, M., Majeed, S., Ban, Y.-E. A., Chen, L., chin Huang, C., Kong, L., Kwon, Y. D., Stuckey, J., Zhou, T., Robinson, J. E., Schief, W. R., Sodroski, J., Wyatt, R., and Kwong, P. D. (2010). Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc Natl Acad Sci U S A*, 107(3):1166–1171.

Pantaleo, G., Graziosi, C., and Fauci, A. S. (1993). New concepts in the immunopathogenesis of human immunodeficiency virus infection. *N Engl J Med*, 328(5):327–335.

Paraskevis, D., Magiorkinis, M., Vandamme, A. M., Kostrikis, L. G., and Hatzakis, A. (2001). Re-analysis of human immunodeficiency virus type 1 isolates from Cyprus and Greece, initially designated 'subtype I', reveals a unique complex A/G/H/K/? mosaic pattern. *J Gen Virol*, 82(Pt 3):575–580.

Parkin, N. T., Chamorro, M., and Varmus, H. E. (1992). Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. *J Virol*, 66(8):5147–5151.

Pastore, C., Nedellec, R., Ramos, A., Pontow, S., Ratner, L., and Mosier, D. E. (2006). Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. *J Virol*, 80(2):750–758.

Pastore, C., Ramos, A., and Mosier, D. E. (2004). Intrinsic obstacles to human immunodeficiency virus type 1 coreceptor switching. *J Virol*, 78(14):7565–7574.

Phillips, A. N. (1996). Reduction of HIV concentration during acute infection: independence from a specific immune response. *Science*, 271(5248):497–499.

Picker, L. J. (2006). Immunopathogenesis of acute AIDS virus infection. *Curr Opin Immunol*, 18(4):399–405.

Pillai, S., Good, B., Richman, D., and Corbeil, J. (2003). A new perspective on V3 phenotype prediction. *AIDS Res Hum Retroviruses*, 19(2):145–149.

Plantier, J.-C., Leoz, M., Dickerson, J. E., Oliveira, F. D., Cordonnier, F., Lemée, V., Damond, F., Robertson, D. L., and Simon, F. (2009). A new human immunodeficiency virus derived from gorillas. *Nat Med*, 15(8):871–872.

Popovic, M., Sarngadharan, M. G., Read, E., and Gallo, R. C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science*, 224(4648):497–500.

Pornillos, O., Ganser-Pornillos, B. K., Kelly, B. N., Hua, Y., Whitby, F. G., Stout, C. D., Sundquist, W. I., Hill, C. P., and Yeager, M. (2009). X-ray structures of the hexameric building block of the HIV capsid. *Cell*, 137(7):1282–1292.

Prosperi, M. C. F., Fanti, I., Ulivi, G., Micarelli, A., Luca, A. D., and Zazzi, M. (2009). Robust Supervised and Unsupervised Statistical Learning for HIV Type 1 Coreceptor Usage Analysis. *AIDS Res Hum Retroviruses*, 25(3):305–314.

Quaranta, M. G., Mattioli, B., Giordani, L., and Viora, M. (2009). Immunoregulatory effects of HIV-1 Nef protein. *Biofactors*, 35(2):169–174.

Rambaut, A., Posada, D., Crandall, K. A., and Holmes, E. C. (2004). The causes and consequences of HIV evolution. *Nat Rev Genet*, 5(1):52–61.

Raymond, S., Delobel, P., Mavigner, M., Cazabat, M., Souyris, C., Encinas, S., Bruel, P., Sandres-Sauné, K., Marchou, B., Massip, P., and Izopet, J. (2010). Development and performance of a new recombinant virus phenotypic entry assay to determine HIV-1 coreceptor usage. *J Clin Virol*, 47(2):126–130.

Raymond, S., Delobel, P., Mavigner, M., Cazabat, M., Souyris, C., Sandres-Sauné, K., Cuzin, L., Marchou, B., Massip, P., and Izopet, J. (2008). Correlation between genotypic predictions based on V3 sequences and phenotypic determination of HIV-1 tropism. *AIDS*.

Reeves, J. D., Gallo, S. A., Ahmad, N., Miamidian, J. L., Harvey, P. E., Sharron, M., Pohlmann, S., Sfakianos, J. N., Derdeyn, C. A., Blumenthal, R., Hunter, E., and Doms, R. W. (2002). Sensitivity of HIV-1 to entry inhibitors correlates with envelope/coreceptor affinity, receptor density, and fusion kinetics. *Proc Natl Acad Sci U S A*, 99(25):16249–16254.

Regoes, R. R. and Bonhoeffer, S. (2005). The HIV coreceptor switch: a population dynamical perspective. *Trends Microbiol*, 13(6):269–277.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., Benenson, M., Gurunathan, S., Tartaglia, J., McNeil, J. G., Francis, D. P., Stablein, D., Birx, D. L., Chunsuttiwat, S., Khamboonruang, C., Thongcharoen, P., Robb, M. L., Michael, N. L., Kunasol, P., Kim, J. H., and Investigators, M. O. P. H.-T. A. V. E. G. (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N Engl J Med*, 361(23):2209–2220.

Resch, W., Hoffman, N., and Swanstrom, R. (2001). Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. *Virology*, 288(1):51–62.

Richman, D. D. and Bozzette, S. A. (1994). The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. *J Infect Dis*, 169(5):968–974.

Richman, D. D., Margolis, D. M., Delaney, M., Greene, W. C., Hazuda, D., and Pomerantz, R. J. (2009). The challenge of finding a cure for HIV infection. *Science*, 323(5919):1304–1307.

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S., and Korber, B. (2000). HIV-1 nomenclature proposal. *Science*, 288(5463):55–56.

Romani, B., Engelbrecht, S., and Glashoff, R. H. (2009). Antiviral roles of APOBEC proteins against HIV-1 and suppression by Vif. *Arch Virol*, 154(10):1579–1588.

Rose, J. D., Rhea, A. M., Weber, J., and Quiñones-Mateu, M. E. (2009). Current tests to evaluate HIV-1 coreceptor tropism. *Curr Opin HIV AIDS*, 4(2):136–142.

Rosen, O., Samson, A. O., and Anglister, J. (2008). Correlated mutations at gp120 positions 322 and 440: implications for gp120 structure. *Proteins*, 71(3):1066–1070.

Rothberg, J. M. and Leamon, J. H. (2008). The development and impact of 454 sequencing. *Nat Biotechnol*, 26(10):1117–1124.

Saag, M., Goodrich, J., Fätkenheuer, G., Clotet, B., Clumeck, N., Sullivan, J., Westby, M., van der Ryst, E., Mayer, H., and Group, A. S. (2009). A double-blind, placebo-controlled trial of maraviroc in treatment-experienced patients infected with non-R5 HIV-1. *J Infect Dis*, 199(11):1638–1647.

Safrit, J. T., Andrews, C. A., Zhu, T., Ho, D. D., and Koup, R. A. (1994). Characterization of human immunodeficiency virus type 1-specific cytotoxic T lymphocyte clones isolated during acute seroconversion: recognition of autologous virus sequences within a conserved immunodominant epitope. *J Exp Med*, 179(2):463–472.

Saksena, N. K., Rodes, B., Wang, B., and Soriano, V. (2007). Elite HIV controllers: myth or reality? *AIDS Rev*, 9(4):195–207.

Samson, M., Labbe, O., Mollereau, C., Vassart, G., and Parmentier, M. (1996a). Molecular cloning and functional expression of a new human CC-chemokine receptor gene. *Biochemistry*, 35(11):3362–3367.

Samson, M., Libert, F., Doranz, B. J., Rucker, J., Liesnard, C., Farber, C. M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C., Muyldermans, G., Verhofstede, C., Burtonboy, G., Georges, M., Imai, T., Rana, S., Yi, Y., Smyth, R. J., Collman, R. G., Doms, R. W., Vassart, G., and Parmentier, M. (1996b). Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature*, 382(6593):722–725.

Sander, O., Sing, T., Sommer, I., Low, A. J., Cheung, P. K., Harrigan, P. R., Lengauer, T., and Domingues, F. S. (2007). Structural Descriptors of gp120 V3 Loop for the Prediction of HIV-1 Coreceptor Usage. *PLoS Comput Biol*, 3(3):e58.

Santiago, M. L., Range, F., Keele, B. F., Li, Y., Bailes, E., Bibollet-Ruche, F., Fruteau, C., Noë, R., Peeters, M., Brookfield, J. F. Y., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (2005). Simian immunodeficiency virus infection in free-ranging sooty mangabeys (Cercocebus atys atys) from the Taï Forest, Côte d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. *J Virol*, 79(19):12515–12527.

Schindler, M., Rajan, D., Banning, C., Wimmer, P., Koppensteiner, H., Iwanski, A., Specht, A., Sauter, D., Dobner, T., and Kirchhoff, F. (2010). Vpu serine 52 dependent counteraction of tetherin is required for HIV-1 replication in macrophages, but not in ex vivo human lymphoid tissue. *Retrovirology*, 7:1.

Schuitemaker, H., Koot, M., Kootstra, N. A., Dercksen, M. W., de Goede, R. E., van Steenwijk, R. P., Lange, J. M., Schattenkerk, J. K., Miedema, F., and Tersmette, M. (1992). Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population. *J Virol*, 66(3):1354–1360.

Shafer, R. W. and Schapiro, J. M. (2008). HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev*, 10(2):67–84.

Shehu-Xhilaga, M., Crowe, S. M., and Mak, J. (2001). Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J Virol*, 75(4):1834–1841.

Simmonds, P., Balfe, P., Ludlam, C. A., Bishop, J. O., and Brown, A. J. (1990). Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol*, 64(12):5840–5850.

Simon, V. and Ho, D. D. (2003). HIV-1 dynamics in vivo: implications for therapy. *Nat Rev Microbiol*, 1(3):181–190.

Simon, V., Ho, D. D., and Karim, Q. A. (2006). HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. *Lancet*, 368(9534):489–504.

Sing, T., Beerenwinkel, N., and Lengauer, T. (2004). Learning Mixtures of Localized Rules by Maximizing the Area Under the ROC Curve. In Hernández-Orallo, J., Ferri, C., Lachiche, N., and Flach, P. A., editors, *1st International Workshop on ROC Analysis in Artificial Intelligence*, pages 89–96, Valencia, Spain.

Sing, T., Low, A. J., Beerenwinkel, N., Sander, O., Cheung, P. K., Domingues, F. S., Büch, J., Däumer, M., Kaiser, R., Lengauer, T., and Harrigan, P. R. (2007). Predicting HIV coreceptor usage on the basis of genetic and clinical covariates. *Antivir Ther*, 12(7):1097–1106.

Sing, T., Svicher, V., Beerenwinkel, N., Ceccherini-Silberstein, F., Däumer, M., Kaiser, R., Walter, H., Korn, K., Hoffmann, D., Oette, M., Rockstroh, J. K., Fätkenheuer, G., Perno, C.-F., and Lengauer, T. (2005). Characterization of Novel HIV Drug Resistance Mutations Using Clustering, Multidimensional Scaling and SVM-Based Feature Ranking. In Jorge, A., Torgo, L., Brazdil, P., Camacho, R., and Gama, J., editors, *9th European Conference on Principles and Practice of Knowledge Discoveryin Databases*, volume 3721 of *Lecture Notes in Computer Science*, pages 285–296. Springer.

Skrabal, K., Low, A. J., Dong, W., Sing, T., Cheung, P. K., Mammano, F., and Harrigan, P. R. (2007). Determining Human Immunodeficiency Virus Coreceptor Use in a Clinical Setting: Degree of Correlation between Two Phenotypic Assays and a Bioinformatic Model. *J Clin Microbiol*, 45(2):279–284.

Smith, K. (2003). The HIV vaccine saga. *Med Immunol*, 2(1):1.

Smith, K. (2005). The continuing HIV vaccine saga: naked emperors alongside fairy godmothers. *Med Immunol*, 4(1):6.

Smith, K. A. (2006). The continuing HIV vaccine saga: is a paradigm shift necessary? *Med Immunol*, 5(1):2.

Smith, P. F., Ogundele, A., Forrest, A., Wilton, J., Salzwedel, K., Doto, J., Allaway, G. P., and Martin, D. E. (2007). Phase I and II study of the safety, virologic effect, and pharmacokinetics/pharmacodynamics of single-dose 3-o-(3',3'-dimethylsuccinyl)betulinic acid (bevirimat) against human immunodeficiency virus infection. *Antimicrob Agents Chemother*, 51(10):3574–3581.

Study, E. H. C. (2006). Long-term non-progressors–international study recruiting. *AIDS Treat News*, page 5.

Swenson, L. C., Boehme, R., Thielen, A., McGovern, R. A., and Harrigan, P. R. (2010a). Genotypic determination of HIV-1 tropism in the clinical setting. *HIV Therapy*, 4:293–303.

Swenson, L. C., Moores, A., Low, A. J., Thielen, A., Dong, W., Woods, C., Jensen, M. A., Wynhoven, B., Chan, D., Glascock, C., and Harrigan, P. R. (2010b). Improved Detection of CXCR4-Using HIV by V3 Genotyping: Application of Population-Based and "Deep" Sequencing to Plasma RNA and Proviral DNA. *J Acquir Immune Defic Syndr*.

Swenson, L. C., Moores, A., Low, A. J., Thielen, A., Dong, W., Woods, C., Jensen, M. A., Wynhoven, B., Chan, D., Glascock, C., and Harrigan, P. R. (2010c). Improved detection of CXCR4-using HIV by V3 genotyping: application of population-based and "deep" sequencing to plasma RNA and proviral DNA. *J Acquir Immune Defic Syndr*, 54(5):506–510.

Team, R. D. C. (2005). R: A language and environment for statistical computing. Technical report, R Foundation for Statistical Computing, Vienna, Austria.

Tersmette, M., Gruters, R. A., de Wolf, F., de Goede, R. E., Lange, J. M., Schellekens, P. T., Goudsmit, J., Huisman, H. G., and Miedema, F. (1989). 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*, 63(5):2118–2125.

Thielen, A., Lengauer, T., Harrigan, P. R., Swenson, L., Dong, W. W., McGovern, R. A., Lewis, M., James, I., Heera, J., and Valdez, H. (2010a). Mutations within GP41 are correlated with coreceptor tropism but do not substantially improve coreceptor usage prediction. In *Reviews in Antiviral Therapy & Infectious Diseases*, volume 2010, pages 47–47, Sorrento, Italy. Virology Education.

Thielen, A., Sichtig, N., Braun, P., Däumer, M., Walter, H., Noah, C., Wolf, E., Müller, H., Stürmer, M., Lengauer, T., Kaiser, R., and Obermeier, M. (2009). Performance of genotypic coreceptor measurement using geno2pheno[coreceptor] in B- and non-B HIV subtypes in a large cohort of therapy-experienced patients in Germany. In *Reviews in Antiviral Therapy*, volume 2009, pages 99–99, Stockholm, Sweden. Virology Education.

Thielen, A., Sichtig, N., Kaiser, R., Lam, J., Harrigan, P. R., and Lengauer, T. (2010b). Improved prediction of HIV?1 coreceptor usage with sequence information from the second hypervariable loop of gp120. *J Infect Dis*, 202(9):1435–1443.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22):4673–4680.

Torres, G. (1995). ACTG 175 and Delta. *GMHC Treat Issues*, 9(10):2–3.

Trouplin, V., Salvatori, F., Cappello, F., Obry, V., Brelot, A., Heveker, N., Alizon, M., Scarlatti, G., Clavel, F., and Mammano, F. (2001). Determination of coreceptor usage of human immunodeficiency virus type 1 from patient plasma samples by using a recombinant phenotypic assay. *J Virol*, 75(1):251–259.

Tscherning, C., Alaeus, A., Fredriksson, R., Björndal, A., Deng, H., Littman, D. R., Fenyö, E. M., and Albert, J. (1998). Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. *Virology*, 241(2):181–188.

UNAIDS (2008). Report on the global HIV/AIDS epidemic 2008. Technical report, UNAIDS.

UNAIDS (2009a). AIDS epidemic update. Technical report, UNAIDS.

UNAIDS, I. T. T. o. H.-r. T. R. (2009b). Mapping of Restrictions on the entry, stay and residence of people living with HIV. Technical report, UNAIDS.

Valdez, H., Lewis, M., Delogne, Simpson, P., Chapman, D., McFadyen, L., Coakley, E., van der Ryst, E., and Westby, M. (2008). Weighted OBT susceptibility score (wOBTss) is a stronger predictor of virologic response at 48 weeks than baseline tropism result in MOTIVATE 1 and 2. In *48th Annual ICAAC/IDSA 46th Annual Meeting. October 25-28, 2008. Washington D.C.*.

Vandekerckhove, L., Wensing, A., Kaiser, R., Brun-Vezinet, F., Clotet, B., De Luca, A., Dressler, S., Garcia, F., Geretti, A., Klimkait, T., Korn, K., Masquelier, B., Perno, C., Schapiro, J., Soriano, V., Sonnerborg, A., Vandamme, A., Verhofstede, C., Walter, H., Zazzi, M., and Boucher, C. (2010). Consensus statement of the European guidelines on clinical management of HIV-1 tropism testing. *Journal of the International AIDS Society*, 13(Suppl 4):O7.

van't Wout, A. B., Kootstra, N. A., Mulder-Kampinga, G. A., van Lent, N. A., Scherpbier, H. J., Veenstra, J., Boer, K., Coutinho, R. A., Miedema, F., and Schuitemaker, H. (1994). Macrophage-tropic variants initiate human immunodeficiency virus type 1 infection after sexual, parenteral, and vertical transmission. *J Clin Invest*, 94(5):2060–2067.

Veazey, R., T. Ketas and, J. D., Klasse, P., and Moore, J. (2010). Protection of Rhesus Macaques from Vaginal Infection by Maraviroc, an Inhibitor of HIV-1 Entry via the CCR5 Co-receptor. In *17th Conference on Retroviruses and Opportunistic Infections*, San Francisco, US.

Verheyen, J., Litau, E., Sing, T., Däumer, M., Balduin, M., Oette, M., Fätkenheuer, G., Rockstroh, J. K., Schuldenzucker, U., Hoffmann, D., Pfister, H., and Kaiser, R. (2006). Compensatory mutations at the HIV cleavage sites p7/p1 and p1/p6-gag in therapy-naive and therapy-experienced patients. *Antivir Ther*, 11(7):879–887.

Wainberg, M. A. and Albert, J. (2010). Can the further clinical development of bevirimat be justified? *AIDS*, 24(5):773–774.

Walker, B. D. and Burton, D. R. (2008). Toward an AIDS vaccine. *Science*, 320(5877):760–764.

Weiss, R. A. (2003). HIV and AIDS in relation to other pandemics. Among the viruses plaguing humans, HIV is a recent acquisition. Its outstanding success as an infection poses immense scientific challenges to human health and raises the question "What comes nest?". *EMBO Rep*, 4 Spec No:S10–S14.

Wensing, A. M. J., van Maarseveen, N. M., and Nijhuis, M. (2010). Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res*, 85(1):59–74.

Westby, M. and van der Ryst, E. (2005). CCR5 antagonists: host-targeted antivirals for the treatment of HIV infection. *Antivir Chem Chemother*, 16(6):339–354.

Westby, M. and van der Ryst, E. (2010). CCR5 antagonists: host-targeted antiviral agents for the treatment of HIV infection, 4 years on. *Antivir Chem Chemother*, 20(5):179–192.

Whitcomb, J. M., Huang, W., Fransen, S., Limoli, K., Toma, J., Wrin, T., Chappey, C., Kiss, L. D. B., Paxinos, E. E., and Petropoulos, C. J. (2007). Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism. *Antimicrob Agents Chemother*, 51(2):566–575.

Whiteside, A., Andrade, C., Arrehag, L., Dlamini, S., Ginindza, T., and Parikh, A. (2006). The Socio-Economic Impact of HIV/AIDS in Swaziland. Technical report, Health Economics and HIV/AIDS Research Division, Durban, South Africa.

Wright, E. J. (2009). Neurological disease: the effects of HIV and antiretroviral therapy and the implications for early antiretroviral therapy initiation. *Curr Opin HIV AIDS*, 4(5):447–452.

Wyatt, R. and Sodroski, J. (1998). The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science*, 280(5371):1884–1888.

# Appendix A: Tables

| *POS* | *HXB2* | *MUT* | *fR5 (%)* | *fX4 (%)* | *LOR* | *95% CI* |
|---|---|---|---|---|---|---|
| 294 | T | A | 1.27 | 6.59 | 0.69 | [1.414, 2.828] |
| 297 | N | F | 0 | 4.4 | N/A | N/A |
| | | Y | 2.54 | 23.08 | 1.25 | [2.440, 5.019] |
| | | S | 11.86 | 2.2 | -2.64 | [0.050, 0.101] |
| 298 | N | E | 0 | 4.4 | N/A | N/A |
| | | N | 97.88 | 69.23 | -1.30 | [0.102, 0.728] |
| | | K | 0 | 4.4 | N/A | N/A |
| | | Q | 0 | 3.3 | N/A | N/A |
| | | Y | 0 | 6.59 | N/A | N/A |
| 299 | N | I | 0 | 5.49 | N/A | N/A |
| | | Y | 0 | 12.09 | N/A | N/A |
| | | K | 0 | 6.59 | N/A | N/A |
| | | N | 98.73 | 69.23 | -1.31 | [0.080, 0.911] |
| 300 | T | T | 97.88 | 65.93 | -1.35 | [0.097, 0.688] |
| | | I | 0.42 | 14.29 | 2.56 | [9.140, 18.48] |
| | | A | 0 | 3.3 | N/A | N/A |
| | | K | 0 | 7.69 | N/A | N/A |
| 301 | R | I | 1.69 | 7.69 | 0.56 | [1.236, 2.477] |
| | | R | 94.92 | 81.32 | -1.11 | [0.151, 0.718] |
| 302 | K | Q | 6.78 | 14.29 | -0.21 | [0.570, 1.157] |
| 303 | R | K | 0.42 | 15.38 | 2.65 | [9.918, 20.10] |
| | | R | 0.42 | 37.36 | 3.54 | [23.52, 50.01] |
| | | S | 88.14 | 24.18 | -2.25 | [0.064, 0.174] |
| 304 | I | T | 2.12 | 16.48 | 1.10 | [2.102, 4.272] |
| 305 | R | S | 2.54 | 16.48 | 0.92 | [1.754, 3.566] |
| | | R | 44.07 | 28.57 | -1.39 | [0.170, 0.366] |
| | | Y | 0.42 | 7.69 | 1.95 | [4.988, 9.990] |
| | | N | 5.08 | 0 | N/A | N/A |
| 306 | I | I | 81.78 | 67.03 | -1.15 | [0.186, 0.534] |
| | | T | 0.42 | 6.59 | 1.80 | [4.278, 8.555] |
| | | R | 0 | 3.3 | N/A | N/A |
| 307 | Q | G | 0 | 9.89 | N/A | N/A |
| | | - | 99.58 | 86.81 | -1.09 | [0.042, 2.644] |
| 308 | R | - | 100 | 96.7 | -0.99 | N/A |
| | | H | 0 | 3.3 | N/A | N/A |
| 308a | - | I | 0 | 9.89 | N/A | N/A |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | - | 100 | 86.81 | -1.09 | N/A |
| 310 | P | P | 97.46 | 68.13 | -1.31 | [0.108, 0.671] |
| | P | Q | 0 | 12.09 | N/A | N/A |
| | P | R | 0 | 6.59 | N/A | N/A |
| | P | L | 0.85 | 7.69 | 1.25 | [2.464, 4.937] |
| 312 | R | R | 33.47 | 57.14 | -0.42 | [0.429, 1.007] |
| | R | Q | 55.93 | 34.07 | -1.45 | [0.157, 0.350] |
| | R | H | 0 | 4.4 | N/A | N/A |
| 313 | A | V | 8.47 | 23.08 | 0.05 | [0.731, 1.509] |
| | | A | 63.14 | 48.35 | -1.22 | [0.192, 0.453] |
| 314 | F | L | 6.36 | 25.27 | 0.43 | [1.064, 2.204] |
| | | F | 83.9 | 52.75 | -1.42 | [0.147, 0.399] |
| | | Y | 1.27 | 6.59 | 0.69 | [1.414, 2.829] |
| | | V | 0 | 5.49 | N/A | N/A |
| 315 | V | H | 2.54 | 10.99 | 0.51 | [1.175, 2.367] |
| | | Y | 88.14 | 74.73 | -1.12 | [0.179, 0.597] |
| 316 | T | A | 63.56 | 27.47 | -1.79 | [0.111, 0.249] |
| | | T | 22.88 | 54.95 | -0.08 | [0.610, 1.404] |
| 317 | I | N | 1.69 | 6.59 | 0.41 | [1.062, 2.126] |
| | | T | 91.53 | 68.13 | -1.25 | [0.154, 0.532] |
| | | S | 0.85 | 7.69 | 1.25 | [2.464, 4.937] |
| 318 | G | R | 0.42 | 4.4 | 1.40 | [2.861, 5.703] |
| | | K | 0.85 | 10.99 | 1.61 | [3.514, 7.073] |
| | | D | 0.42 | 4.4 | 1.40 | [2.861, 5.703] |
| | | E | 1.27 | 7.69 | 0.85 | [1.649, 3.304] |
| | | G | 82.2 | 54.95 | -1.36 | [0.157, 0.422] |
| 319 | K | K | 1.69 | 24.18 | 1.71 | [3.843, 7.919] |
| | | D | 47.88 | 12.09 | -2.33 | [0.067, 0.141] |
| | | R | 1.69 | 15.38 | 1.26 | [2.464, 4.996] |
| | | - | 0 | 5.49 | N/A | N/A |
| | | I | 93.64 | 73.63 | -1.19 | [0.152, 0.603] |
| 320 | I | K | 0 | 6.59 | N/A | N/A |
| | | V | 2.97 | 9.89 | 0.25 | [0.905, 1.821] |
| | | I | 91.53 | 67.03 | -1.26 | [0.152, 0.522] |
| 321 | G | G | 98.73 | 92.31 | -1.02 | [0.091, 1.425] |
| 322 | N | D | 83.47 | 71.43 | -1.11 | [0.189, 0.573] |
| 323 | M | I | 97.88 | 86.81 | -1.07 | [0.117, 0.997] |
| 324 | R | G | 0 | 3.3 | N/A | N/A |
| | R | R | 99.15 | 93.41 | -1.01 | [0.072, 1.829] |

Table 1: Significant mutations within V3. POS: position in HXB2, HXB2: residue at this position in HXB2, MUT: residue tested for association with phenotype, fR5 & fX4: frequencies in R5- and X4-viruses, respectively, LOR: Log odds ratio, 95% CI: 95% confidence intervals

| *dataset I* | *dataset II* | *feature* | *mr5-I* | *mx4-I* | *mr5-II* | *mx4-II* | *pr5* | *px4* |
|---|---|---|---|---|---|---|---|---|
| *GERMANY* | *HOMER* | length | 34.83 | 34.62 | 34.92 | 34.86 | -3.64 | -1.57 |
| *GERMANY* | *HOMER* | hydrophobic | 21.10 | 21.00 | 21.53 | 21.54 | -10.24 | -4.58 |
| *GERMANY* | *HOMER* | positive | 6.55 | 7.58 | 6.76 | 7.33 | -5.91 | -1.46 |
| *GERMANY* | *HOMER* | charge | 3.38 | 4.55 | 3.27 | 4.09 | -1.60 | -3.50 |
| *GERMANY* | *L.A.* | length | 34.83 | 34.62 | 34.71 | 34.48 | -1.35 | -2.43 |
| *GERMANY* | *L.A.* | hydrophobic | 21.10 | 21.00 | 20.67 | 20.62 | -9.15 | -2.26 |
| *GERMANY* | *L.A.* | small | 16.02 | 15.48 | 15.80 | 14.67 | -3.15 | -6.49 |
| *GERMANY* | *L.A.* | proline | 2.11 | 2.07 | 2.05 | 1.87 | -1.87 | -4.66 |
| *GERMANY* | *L.A.* | tiny | 8.15 | 7.69 | 7.72 | 6.82 | -12.35 | -12.38 |
| *GERMANY* | *L.A.* | aromatic | 3.57 | 3.62 | 3.29 | 3.47 | $<$ -50 | -1.80 |
| *GERMANY* | *L.A.* | charge | 3.38 | 4.55 | 3.19 | 4.92 | -4.75 | -2.84 |
| *GERMANY* | *MOTIVATE* | small | 16.02 | 15.48 | 16.20 | 16.06 | -1.70 | -5.90 |
| *GERMANY* | *MOTIVATE* | tiny | 8.15 | 7.69 | 8.37 | 8.06 | -3.04 | -3.16 |
| *GERMANY* | *MOTIVATE* | positive | 6.55 | 7.58 | 6.67 | 7.88 | -1.56 | -2.58 |
| *HOMER* | *GERMANY* | positive | 6.76 | 7.33 | 6.55 | 7.58 | -5.91 | -1.46 |
| *HOMER* | *GERMANY* | charge | 3.27 | 4.09 | 3.38 | 4.55 | -1.60 | -3.50 |
| *HOMER* | *ITALY* | hydrophobic | 21.53 | 21.54 | 20.97 | 21.06 | -9.64 | -2.83 |
| *HOMER* | *ITALY* | aliphatic | 5.18 | 5.59 | 5.06 | 5.19 | -2.13 | -3.53 |
| *HOMER* | *L.A.* | length | 34.92 | 34.86 | 34.71 | 34.48 | -9.06 | -5.24 |
| *HOMER* | *L.A.* | hydrophobic | 21.53 | 21.54 | 20.67 | 20.62 | $<$ -50 | -10.93 |
| *HOMER* | *L.A.* | small | 15.57 | 15.74 | 15.80 | 14.67 | -4.02 | -7.16 |
| *HOMER* | *L.A.* | tiny | 8.19 | 8.09 | 7.72 | 6.82 | $<$ -50 | $<$ -50 |
| *HOMER* | *L.A.* | aromatic | 3.75 | 3.70 | 3.29 | 3.47 | $<$ -50 | -3.36 |
| *HOMER* | *L.A.* | negative | 1.81 | 1.69 | 1.66 | 1.33 | -6.20 | -5.61 |
| *HOMER* | *L.A.* | glyco | 0.99 | 0.92 | 0.97 | 0.66 | -3.18 | -8.65 |
| *HOMER* | *L.A.* | charge | 3.27 | 4.09 | 3.19 | 4.92 | -1.64 | -8.17 |
| *HOMER* | *MOTIVATE* | polar | 16.40 | 16.74 | 16.59 | 17.17 | -4.78 | -4.00 |
| *HOMER* | *MOTIVATE* | small | 15.57 | 15.74 | 16.20 | 16.06 | -22.21 | -2.24 |
| *HOMER* | *MOTIVATE* | positive | 6.76 | 7.33 | 6.67 | 7.88 | -2.85 | -5.44 |
| *HOMER* | *MOTIVATE* | charged | 7.29 | 7.87 | 7.03 | 8.29 | -9.57 | -2.73 |
| *HOMER* | *MOTIVATE* | charge | 3.27 | 4.09 | 3.42 | 4.73 | -3.24 | -7.05 |
| *ITALY* | *L.A.* | hydrophobic | 20.97 | 21.06 | 20.67 | 20.62 | -4.28 | -3.25 |
| *ITALY* | *L.A.* | small | 16.00 | 15.81 | 15.80 | 14.67 | -2.79 | -9.24 |
| *ITALY* | *L.A.* | proline | 2.12 | 2.10 | 2.05 | 1.87 | -2.28 | -4.10 |
| *ITALY* | *L.A.* | tiny | 8.08 | 7.86 | 7.72 | 6.82 | -5.96 | -12.56 |
| *ITALY* | *L.A.* | charge | 3.36 | 4.54 | 3.19 | 4.92 | -3.17 | -2.13 |
| *ITALY* | *MOTIVATE* | hydrophobic | 20.97 | 21.06 | 21.24 | 21.39 | -1.88 | -1.40 |
| *ITALY* | *MOTIVATE* | positive | 6.49 | 7.39 | 6.67 | 7.88 | -1.94 | -3.36 |
| *L.A.* | *MOTIVATE* | length | 34.71 | 34.48 | 34.88 | 34.82 | -3.93 | -8.54 |
| *L.A.* | *MOTIVATE* | hydrophobic | 20.67 | 20.62 | 21.24 | 21.39 | -23.08 | -10.53 |
| *L.A.* | *MOTIVATE* | polar | 16.41 | 16.59 | 16.59 | 17.17 | -3.49 | -6.29 |
| *L.A.* | *MOTIVATE* | small | 15.80 | 14.67 | 16.20 | 16.06 | -11.27 | -21.55 |
| *L.A.* | *MOTIVATE* | proline | 2.05 | 1.87 | 2.12 | 2.10 | -3.49 | -7.96 |
| *L.A.* | *MOTIVATE* | tiny | 7.72 | 6.82 | 8.37 | 8.06 | -36.10 | -25.39 |
| *L.A.* | *MOTIVATE* | aromatic | 3.29 | 3.47 | 3.60 | 3.63 | -30.54 | -2.18 |
| *L.A.* | *MOTIVATE* | positive | 5.97 | 7.38 | 6.67 | 7.88 | -58.26 | -3.40 |
| *L.A.* | *MOTIVATE* | negative | 1.66 | 1.33 | 1.74 | 1.65 | -2.46 | -6.01 |
| *L.A.* | *MOTIVATE* | charged | 6.25 | 7.66 | 7.03 | 8.29 | -51.12 | -5.27 |

Table 2: Differences in position-independent properties between two datasets. mr5-I (mr5-II) depicts the average value in R5-samples of dataset I (II), mx4-I and II the respective average values in X4-samples. pr5 and px4 are the corresponding log10 p-values.

# Appendix B: List of Publications

Christian Pou, Francisco M. Codoñer, **Alexander Thielen**, Rocío Bellido, Susana Pérez-Álvarez, Cecilia Cabrera, Judith Dalmau, Marta Curriu, Yolanda Lie, Marc Noguera, Mattia Schiaulini, Jordi Puig, Javier Martínez-Picado, Juliá Blanco, Eoin Coackley, Martin Däumer, Bonaventura Clotet, Roger Paredes (submitted). HIV-1 Tropism Kinetics during Suppressive First-Line Antiretroviral Therapy: Diagnostic Accuracy of Genotypic and Phenotypic Assays. *Clinical Infectious Diseases (submitted)*

Francisco M Codoñer, Christian Pou, **Alexander Thielen**, Federico García, Rafael Delgado, David Dalmau, Miguel Álvarez-Tejado, Lidia Ruiz, Bonaventura Clotet, Roger Paredes (submitted). Added Value of Deep HIV-1 Sequencing in Deep Salvage Therapy. *PLoS One (submitted)*

Martin Stürmer, Thomas Berg, Patrick Braun, Martin Däumer, Josef Eberle, Robert Ehret, Rolf Kaiser, Nils Kleinkauf, Klaus Korn, Claudia Kücherer, Harm Müller, Martin Obermeier, **Alexander Thielen**, Eva Wolf, Hauke Walter (submitted). HIV-GRADE, a Publicly Available Internet-based Algorithm for the Interpretation of HIV-1 Genotypic Drug Resistance to Antiretroviral Drugs. *Antiviral Therapy (subtmited)*

Martin Däumer, Rolf Kaiser, Rolf Klein, Thomas Lengauer, Bernhard Thiele, **Alexander Thielen** (submitted). Genotypic tropism testing by massively parallel sequencing: qualitative and quantitative analysis. *BMC Medical Informatics and Decision Making (subtmited)*

**Alexander Thielen**, Thomas Lengauer, Luke C Swenson, Winnie W.Y. Dong, Rachel A McGovern, Marilyn Lewis, Ian James, Jayvant Heera, Hernan Valdez, P Richard Harrigan (accepted). Mutations in gp41 are correlated with coreceptor tropism but do not improve prediction methods substantially. *Antiviral Therapy (accepted)*

Stefanie A. Knoepfel, Martin Däumer, **Alexander Thielen**, Karin J. Metzner (2011). In-depth analysis of G-to-A hypermutation rate in HIV-1 DNA induced by endogenous APOBEC3 proteins using massively parallel sequencing. *Journal of Virological Methods 2011, 171, 329-338*

**Alexander Thielen**, Nadine Sichtig, Rolf Kaiser, Jeffrey Lam, P. Richard Harrigan, Thomas Lengauer (2010) Improved Prediction of HIV-1 Coreceptor Usage with Sequence Information from the Second Hypervariable Loop of gp120. *Journal of Infectious Diseases 2010:202*

Francisco M Codoñer, Christian Pou, **Alexander Thielen**, Federico García, Rafael Delgado, David Dalmau, José Ramon Santos, Maria José Buzón, Javier Martínez-Picado,

Miguel Álvarez-Tejado, Bonaventura Clotet, Lidia Ruiz, Roger Paredes (2010). Dynamic Escape of Pre-Existing Raltegravir-Resistant HIV-1 from Raltegravir Selection Pressure. *Antiviral Research 2010, 88, 281-286.*

Rachel A McGovern, **Alexander Thielen**, Theresa Mo, Winnie Dong, Conan K Woods, Douglass Chapman, Marilyn Lewis, Ian James, Jayvant Heera, Hernan Valdez, P Richard Harrigan (2010). Population-based V3 genotypic tropism assay: a retrospective analysis using screening samples from the A4001029 and MOTIVATE studies. *AIDS 2010, 24, 2517-2525.*

Luke C Swenson, Andrew Moores, Andrew J Low, **Alexander Thielen**, Winnie Dong, Conan Woods, Mark A Jensen, Brian Wynhoven, Dennison Chan, Christopher Glascock, P Richard Harrigan (2010). Improved Detection of CXCR4-Using HIV by V3 Genotyping: Application of Population-Based and "Deep" Sequencing to Plasma RNA and Proviral DNA. *J Acquir Immune Defic Syndr. 2010 May 27.*

Thomas Lengauer, André Altmann, **Alexander Thielen**, Rolf Kaiser (2010). Chasing the AIDS virus. *Communications of the ACM, 2010, 53(3): 66-74.*

Christel Kamp, Timo Wolf, Ignacio G Bravo, Benjamin Kraus, Birgit Krause, Britta Neumann, Gudrun Winskowsky, **Alexander Thielen**, Albrecht Werner, Barbara S Schnierle (2010). Decreased HIV diversity after allogeneic stem cell transplantation of an HIV-1 infected patient: a case report. *Virology Journal 2010, 7:55*

Luke C Swenson, Richard Boehme, **Alexander Thielen**, Rachel A McGovern, P Richard Harrigan (2010). Genotypic determination of HIV-1 tropism in the clinical setting. *HIV Therapy 2010; 4(3): 293-303.*

Thomas Lengauer, André Altmann, **Alexander Thielen** (2009). Bioinformatische Unterstützung der Auswahl von HIV-Therapien. *Informatik-Spektrum, 2009, 32(4): 320-331 (in german).*

Katarzyna Bozek, **Alexander Thielen**, Saleta Sierra, Rolf Kaiser, Thomas Lengauer (2009). V3 loop sequence space analysis suggests different evolutionary patterns of CCR5- and CXCR4-tropic HIV. *PLoS One. 2009 Oct 9;4(10):e7387.*

Saleta Sierra, Rolf Kaiser, **Alexander Thielen**, Oliver Sander, Thomas Lengauer (2008). Genotypic Analysis of HIV Co-Receptor Usage. *Entry Inhibitoren - Neue Formen der HIV-Therapie, Springer Medizin, 2008, 31-39 (in german)*

Thomas Lengauer, Oliver Sander, Saleta Sierra, **Alexander Thielen**, Rolf Kaiser (2007). Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol. 2007, 25, 1407-1410.*

Saleta Sierra, Rolf Kaiser, **Alexander Thielen**, Thomas Lengauer (2007). Genotypic coreceptor analysis. *Eur J Med Res. 2007 Oct 15;12(9):453-62.*