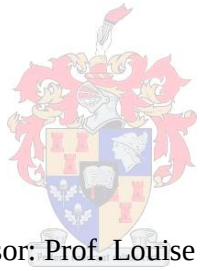
Effect of genetic variants in genes encoding two nuclear receptors (PXR and CAR) on efavirenz levels and treatment outcome in South African HIV-infected females.

by

Enid Nieuwoudt

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Genetics at Stellenbosch University



Supervisor: Prof. Louise Warnich

December 2014

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived, are those of the author and are not necessarily to be attributed to the NRF.

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Enid Nieuwoudt

Date: 2014/11/21

Abstract

Efavirenz is an antiretroviral drug used in the treatment of HIV-positive patients as part of first line triple-highly active antiretroviral therapy. Treatment response varies among individuals and adverse drug reactions tend to occur, as a result of the variation in the rate of efavirenz metabolism among individuals. This is partly caused by genetic variation; therefore the study of genes involved in the metabolism of efavirenz, such as *CYP2B6*, could potentially enhance treatment success. The effect of *CYP2B6* SNP 516G>T (part of the *CYP2B6*6* allele) is particularly important, as individuals homozygous for the minor allele of this SNP have significantly increased efavirenz levels. Furthermore, nuclear receptors, specifically constitutive androstane receptor, encoded by *NR1I3*, and pregnane X receptor, encoded by *NR1I2*, are involved in the regulation of the genes responsible for efavirenz metabolism and could therefore indirectly influence the pharmacokinetics of efavirenz.

The current study identified variants in the *NR1I3* and *NR1I2* genes through *in silico* analysis, bi-directional sequencing and literature searches. A total of nine *NR1I3* and ten *NR1I2* target variants were subsequently genotyped in 132 HIV-positive female patients from the Xhosa and Cape Mixed Ancestry populations. The resulting genotype and allele frequencies were statistically analysed to search for correlations between genetic variations and available efavirenz levels in hair samples, treatment outcome as measured by viral load, and the occurrence of adverse drug reactions. The minor allele of a *NR1I2* 5'-upstream SNP, rs1523128 (6334A>G), was significantly associated with decreased efavirenz levels. From analysis of the effect of composite SNPs, *NR1I3* 5'-upstream SNP rs55802895 (258G>A) in conjunction with *CYP2B6*6*, was significantly associated with efavirenz-levels. It was found that the minor allele of rs55802895 inhibited the effect of *CYP2B6*6*, resulting in normal efavirenz levels for individuals homozygous for the minor allele of both SNPs. Additionally, when the target *NR1I3* and *NR1I2* variants were analysed in conjunction with six SNPs from *CYP1A2*, *CYP2A6*, *CYP3A4* and *CYP3A5*, 11 compound genotypes were shown to be statistically associated with mean EFV plasma levels. The study emphasises the complexity of efavirenz metabolism, and the importance of transcriptional regulation in xenobiotic metabolism.

Opsomming

Efavirenz is ‘n antiretrovirale middel wat gebruik word in die behandeling van HIV-positiewe pasiënte as deel van drievoudige hoogs-aktiewe antiretrovirale terapie. Reaksie op behandeling verskil tussen individue en nadelige newe-effekte, wat veroorsaak word deur die verskil in tempo waarteen efavirenz gemetaboliseer word, neig om voor te kom. Hierdie verskille word gedeeltelik veroorsaak deur genetiese variasie; dus kan die studie van gene betrokke by die metabolisme van efavirenz, soos *CYP2B6*, moontlik die sukses van behandeling verhoog. Die effek van *CYP2B6* SNP 516G>T (deel van die *CYP2B6*6*-alleel) is veral belangrik, want individue wat homosigoties is vir die minderheids-alleel het betekenisvol hoë efavirenz-vlakke. Nukleêre reseptore, spesifiek konstitutiewe androstane reseptor, deur *NR1I3* gekodeer, en pregnane X reseptor, deur *NR1I2* gekodeer, is betrokke by die regulering van die gene verantwoordelik vir efavirenz-metabolisme en kan dus die farmakokinetika van efavirenz beïnvloed.

Die huidige studie het variante in *NR1I3* en *NR1I2* identifiseer deur *in silico*-analise, bi-direksionele volgordebepaling en ’n literatuurstudie. Nege *NR1I3* en tien *NR1I2*-variante in totaal is vervolglik gegenotipeer in 132 HIV-positiewe vroulike pasiënte van Xhosa en Kaapse Gemengde Afkoms populasies. Die gevolglike genotipe- en alleelfrekwensies is statisties geanaliseer om vir korrelasies tussen genetiese variasies en beskikbare efavirenz-vlakke in haarmonsters, uitkoms van behandeling gemeet in virale lading en die voorkoms van nadelige newe-effekte te soek. Daar is gevind dat die minderheids-alleel van ’n *NR1I2* 5’-stroomop SNP, rs1523128 (6334A>G), betekenisvol geassosieer is met ’n daling in efavirenz-vlakke. Vanuit die saamgestelde SNPs, is die *NR1I3* 5’-stroomop SNP rs55802895 (258G>A), tesame met *CYP2B6*6*, betekenisvol geassosieer met efavirenz-vlakke. Daar is gevind dat die minderheids-alleel van rs55802895 die effek van *CYP2B6*6* demp, en gevolglik normale efavirenz-vlakke in individue homosigoties vir die minderheids-allele van albei SNPs veroorsaak. Addisioneel is die teiken *NR1I3* en *NR1I2* variante gemeenskaplik met ses SNPs van *CYP1A2*, *CYP2A6*, *CYP3A4* en *CYP3A5* geanaliseer en 11 gekombineerde genotipes is statisties geassosieer met gemiddelde EFV plasma vlakke. Hierdie studie beklemtoon die kompleksiteit van efavirenz-metabolisme en die belangrikheid van transkripsionele regulering in xenobiotiese metabolisme.

Acknowledgements

I would like to thank the following people and institutions for their much appreciated contributions to my thesis:

Professor Louise Warnich, my supervisor, for accommodating me in the HIV/ARV pharmacogenetics research group, for guidance, wise words and advice,

Patients of TC Newman HIV clinic, Paarl, for participation in this study,

Dr Nelis Grobbelaar and staff of TC Newman HIV Clinic, for help with patient recruitment and collection of clinical data,

Carola Röhrich, for the collection of patient DNA and hair samples, for assistance with the collection of clinical data and for supplying *CYP2B6* data for use in this study,

Prof Monica Gandhi and Dr Yong Huang from the Department of Clinical Pharmacy, University of California, San Francisco, for determining the efavirenz levels in hair samples,

Mr Gustav Victor, for supplying *CYP1A2, CYP2A6, CYP3A4* and *CYP3A5* data for use in this study,

Mrs Lundi Korkie, our laboratory manager, for assistance with genotyping SNP rs3003596 and for motivation and support,

Dr Lize van der Merwe, for statistical analysis,

Dr Natalie Roetz, for help with statistical analysis, editing and guidance,

Dr Britt Drögemöller and Dr Galen Wright, for guidance,

Postgraduate students from Lab 231, for motivation and support,

My parents, John and Sadie Nieuwoudt, for unconditional love and motivation, and financial support,

My siblings, Jaybé and Sadie-Loraine Nieuwoudt, for their love and support,

My friends, for their love and support,

Jesus Christ, for strength, unconditional love and mercy,

The Medical Research Council for financial support, and

The National Research Foundation for a bursary.

Once again, thank you all.

Table of contents

List of tables

List of figures

List of abbreviations and symbols

α	alpha
χ^2	chi-square
°C	degrees Celsius
5’	five-prime
<	less than
µg	micro gram
µl	micro litre
-	minus
>	more than
/	or / per
%	percentage
+	plus
®	registered trade mark
3’	three-prime
x	times
3TC	lamivudine
A	adenine / alanine
AA	African American
aa	amino acid
ABCB	ATP-binding cassette subfamily B gene
ADRs	adverse drug reactions
AF-1	transactivation domain 1
AF-2	transactivation domain 2
Afr	African
AIDS	acquired immunodeficiency syndrome
AKR	aldo-keto reductase
ART	antiretroviral therapy
ARV	antiretroviral
Asn	Asian
ASW	African American population of South West USA
AZT	zidovudine
BHIVA	British HIV Association
BMI	body mass index
bp	base pairs
BSA	bovine serum albumin
C	cytosine
CAF	Central Analytical Facility
CAR	constitutive androstane receptor
Cau	Caucasian
C. briggsae	*Caenorhabditis briggsae*
C. elegans	*Caenorhabditis elegans*
CD4	cluster of differentiation 4
CES	carboxylesterase
CEU	Caucasian population of Europe
CHB	Chinese Han population of Beijing
CMA	Cape Mixed Ancestry
CNS	central nervous system
CRI	co-receptor inhibitor
CSF	cerebrospinal fluid

CYP	cytochrome P450
CYP1A2	Cytochrome P450 subfamily 1A2 gene
CYP2A6	Cytochrome P450 subfamily 2A6 gene
CYP2B6	Cytochrome P450 subfamily 2B6 gene
CYP2C19	Cytochrome P450 subfamily 2C19 gene
CYP2C9	Cytochrome P450 subfamily 2C9 gene
CYP3A4	Cytochrome P450 subfamily 3A4 gene
CYP3A5	Cytochrome P450 subfamily 3A5 gene
CYP4F12	Cytochrome P450 subfamily 4F12 gene
D	aspartic acid
d4T	stavudine
DBD	DNA binding domain
DHHS	Department of Health and Human Services USA
D. melanogaster	*Drosophila melanogaster*
DNA	deoxy-ribonucleic acid
dNTP	deoxynucleotide triphosphate
EDTA	ethylenediaminetetraacetic acid
EFV	efavirenz
et al.	*et alia*, and others
FI	fusion inhibitor
FTC	emtricitabine
FXR	farnesoid X receptor
G	guanine / glycine
GR	glucocorticoid receptor
GST	glutathione-S-transferase
H	histidine
HBV	hepatitis B virus
HIV	human immunodeficiency virus
HNF-α	hepatic nuclear factor alpha
hr	hour
hrs	hours
HWE	Hardy-Weinberg equilibrium
IC_{50}	inhibitory concentration 50
ID	identification
IDT	Integrated DNA Technologies Inc.
IDV	indinavir
Inc.	incorporated
INSTI	integrase strand transfer inhibitor
IQR	inter-quartile range
IRIS	immune reconstitution inflammatory syndrome
kb	kilo base pair
L	leucine
LBD	ligand binding domain
LC	liquid chromatography
LD	linkage disequilibrium
LPV	lopinavir
LS	locus specific
Ltd.	limited
M	methionine
mA	milli-ampere
MAF	minor allele frequency

mg	milligram
$MgCl_2$	magnesium chloride
min	minute / minutes
miRNA	micro-RNA
ml	millilitre
mM	millimolar
mRNA	messenger RNA
MS	mass spectrometry
N	asparagine
n/a	not applicable
N-K	Niger-Kordofanian
NEB	New England Biolabs®, Inc.
NFV	nelfinavir
ng	nanogram
NH_4	ammonium
NLS	nested locus specific
nm	nanometre
NR	nuclear receptor
NR1I2	nuclear receptor subfamily 1I2 gene
NR1I3	nuclear receptor subfamily 1I3 gene
NRTI	nucleoside reverse transcriptase inhibitor
NNRTI	non-nucleoside reverse transcriptase inhibitor
nt	nucleotide
NtRTI	nucleotide reverse transcriptase inhibitor
NVP	nevirapine
OC	oropharyngeal candidiasis
P	probability value
P	proline
P-gp	p-glycoprotein
pBLAST	primer-BLAST
PBREM	phenobarbital-responsive enhancer module
PCR	polymerase chain reaction
PCR-RFLP	polymerase chain reaction-restriction fragment length polymorphism
PI	protease inhibitor
pmol	picomole
Pty.	proprietary limited
PXR	pregnane X receptor
Q	glutamine
R	arginine
RAL	raltegravir
RNA	ribonucleic acid
RT	reverse transcriptase
RTV	ritonavir
RXR	retinoid X receptor
S	serine
SDS	sodium dodecyl sulphate
sec	seconds
SNP	single nucleotide polymorphism
SQV	saquinavir
ss	single strand
SULT	sulfotransferase

T	thymine / threonine
T_A	annealing temperature
TB	tuberculosis
TBE	Tris, boric acid and EDTA
TDF	tenofovir disoproxil fumarate
TDM	therapeutic drug monitoring
TF	transcription factor
TFBS	transcription factor binding site
T_M	melting temperature
^{TM}	trade mark
TSP	temperature switch PCR
u	units
UA	upstream amplicons
UCSF	University of California, San Francisco
UGT	UDP glucuronosyltransferase
UGT1A1	UDP glucuronosyltransferase subfamily 1A1 gene
UGT1A4	UDP glucuronosyltransferase subfamily 1A4 protein
UGT1A9	UDP glucuronosyltransferase subfamily 1A9 protein
UGT2B7	UDP glucuronosyltransferase subfamily 2B7 gene
UK	United Kingdom
UNAIDS	United Nations Programme on HIV/AIDS
USA	United States of America
UTR	untranslated region
UV	ultraviolet
V	valine / volts
v	version
VL	viral load
W	tryptophan
WHO	World Health Organization
w/v	weight per volume
Xh	Xhosa
XRE	xenobiotic response element
YRI	Yoruban population of Nigeria
ZA	South Africa

CHAPTER 1

1. Introduction

Access to antiretroviral therapy (ART) has vastly improved over a period of ten years. ART is now accessible to an increased number of people in 2014 (12.9 million) compared to 2003 (400,000) (World Health Organization, 2004; World Health Organization, 2014). This has greatly increased the problem of inter-individual variation in patient response to antiretroviral (ARV) drugs, and the difficulty of measuring this variation through viral load (VL), cluster of differentiation 4 (CD4) cell count and clinical outcome.

Alternative strategies including therapeutic drug monitoring (TDM) and pharmacogenetic analysis can aid in measuring and managing individual response to treatment through the alteration of ARV dosage. These two methods have already been proven to play a significant role in studies analysing treatment response (Marzolini *et al*., 2001; Fellay *et al*., 2002; Haas *et al*., 2004). TDM is the measurement and monitoring of ARV levels in the body in order to evaluate patient treatment response. ARV levels in the body indicate the rate at which the ARV is broken down. Sub-optimal ARV levels, the possible result of a fast metabolism rate, can cause a negative treatment response, as indicated by persistently high VLs and low CD4 counts. Super-optimal ARV levels, the possible result of a slow metabolism rate, can cause a high occurrence of adverse drug reactions (ADRs).

Efavirenz (EFV), an ARV drug classified as a non-nucleoside reverse transcriptase inhibitor (NNRTI), is prescribed as part of a standardised, first-line, triple-ART regimen in South Africa (National Department of Health, 2010). The inter-individual differences in HIV-positive patient response to EFV treatment can in part be attributed to the variable rate of EFV metabolism among patients, and subsequently, variable long-term EFV levels present in the bodies of different patients (Haas *et al*., 2004). This drug is a significant case to consider, because of its narrow therapeutic range and vast inter-individual pharmacokinetic variation. EFV plasma levels have been associated with VL parameters and the occurrence of EFV-specific ADRs (Marzolini *et al*. 2001; Figueroa *et al*., 2010). Importantly *Cytochrome P450* (*CYP*) *2B6* SNPs *516G>T* (part of the *CYP2B6*6* allele) and *983T>C* (part of the *CYP2B6*16* and **18* alleles) have consistently been associated with an increase in EFV levels (Haas *et al*., 2004; Gandhi *et al*., 2012; Ngaimisi *et al*., 2013).

EFV metabolism is a complex process involving not only the enzymes (CYP2B6, CYP3A, CYP1A2, CYP2A6, and UGT2B7) directly responsible for EFV breakdown into its metabolites, but also the nuclear receptors responsible for the transcriptional regulation of the genes encoding the metabolisers (constitutive androstane receptor, CAR and pregnane X receptor, PXR). These two nuclear receptors play a role in the transcriptional regulation of genes responsible for xenobiotic metabolism, among others. The activity of these two nuclear receptors is, in turn, activated by a range of ligands, of which xenobiotic compounds are the majority. Genetic variants in *NR1I3* and *NR1I2*, encoding for CAR and PXR respectively, have been associated with altered regulatory activity of CAR and PXR (Zhang *et al*., 2001; Ikeda *et al*., 2005), as well as altered EFV levels and the occurrence of EFV-specific ADRs (Wyen *et al*., 2011; Swart *et al*., 2012).

It would therefore be advantageous to investigate whether *NR1I3* and *NR1I2* genotypes affect EFV levels and treatment outcome in South African HIV patients.

CHAPTER 2

2. Literature Review

2.1. Antiretroviral therapy and treatment outcome

2.1.1. Antiretroviral accessibility and management on an international scale

Since the official implementation of the first effective ART in 1996, many HIV/AIDS-infected individuals have experienced significant improvements in both their life expectancy and quality of life (Palella *et al*., 2003; World Health Organization & UNAIDS, 2003). However, most of the patients who received ART were from first world countries, primarily because ARV drugs were more easily available and affordable in these resource-rich, developed countries (World Health Organization, 2004).

In 2003, an estimated 400,000, out of six million patients in developing countries, were receiving ART; 100,000 of these patients were from African countries (World Health Organization, 2004, 2006a). The World Health Organization (WHO) and the United Nations Programme on HIV/AIDS (UNAIDS) thus formed the "3 by 5" Initiative to address the huge demand for ARV drugs in developing countries (World Health Organization, 2004). This joint venture worked towards making ART available for three million people by the year 2005, through the creation of documents to provide healthcare departments, particularly in developing, resource-poor countries, with guidelines on how to implement national ART programmes (World Health Organization, 2004).

Even though the target of providing access to ART for three million people was not achieved by 2005, the number of people on ART tripled from 2003 to the end of 2005 (approximately 1.3 million patients; of which 800,000 were from Africa) (World Health Organization & UNAIDS, 2006). The "3 by 5" Initiative was a success in that it highlighted the huge demand, as well as the limited access to ART, worldwide. Today the WHO, UNAIDS and other important organisations continue to provide guidance for the management of ART in all countries.

The latest global HIV statistics revealed that more than 35 million people are HIV positive, with 70% residing in sub-Saharan Africa (World Health Organization, 2013b; World Health Organization, 2014). A decade after the launch of the 3 by 5 Initiative, approximately 12.9 million people in developing countries are being treated with ARV drugs (World Health Organization, 2014). The number of people eligible for ART currently stands at 17.6 million

according to the WHO 2010 guidelines on antiretroviral therapy and will be extended to 28 million, pending the implementation of the latest recommendations made in 2013 by the WHO (World Health Organization, 2010a; World Health Organization, 2013b; Stover *et al*., 2014; World Health Organization, 2014). By the end of 2013, two million new patients have started receiving ART, the highest number to date, signifying that the rate of access to ART therapy is growing rapidly (World Health Organization, 2014). With the latest revised guidelines released in 2013, the goal of 15 million people receiving ART by the end of 2015 does appear to be achievable (World Health Organization, 2013b; Stover *et al*., 2014).

The WHO ARV guidelines recommend using WHO clinical stage and the CD4 cell count of patients to determine eligibility for ART. The WHO clinical stages represent a classification system for the severity of symptoms associated with AIDS, with the least severe at stage I (nearly asymptomatic) and the most severe at stage IV (critical bacterial, viral and/or fungal infection) (World Health Organization, 2004).

Table 1 lists the current recommended criteria for asymptomatic and symptomatic HIV-positive patients that are eligible for ART, in comparison to previous criteria. These criteria are periodically updated as new evidence from clinical studies emerges, ART coverage goals are reached and new goals, with regards to ART coverage, are set. Compared to the 2006 guideline, the CD4 cut-off value is less stringent in the 2010 guidelines (≤ 350 cells/ml), and even more so in the 2013 edition that is currently being implemented (≤ 500 cells/ml) (World Health Organization, 2010a; 2013b; World Health Organization, 2014). These instructions are in agreement with guidelines from the United States of America (USA) Department of Health and Human Services (DHHS) (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014). Patients with tuberculosis (TB)- or hepatitis B virus (HBV)-infections combined with HIV infections are automatically enlisted for ART (World Health Organization, 2013b).

Table 1: WHO criteria for eligibility for ART, adapted from (World Health Organization, 2010a).

Target group	2013 ART guidelines	2010 ART guidelines	2006 ART guidelines
Asymptomatic patients (stage I)	• CD4 ≤ 500 cells/ml	• CD4 ≤ 350 cells/ml	• CD4 ≤ 200 cells/ml
Symptomatic patients	• CD4 ≤ 500 cells/ml • Priority given to: – Stage III or IV – Or CD4 count ≤ 350 cells/ml	• Stage II with CD4 ≤ 350 cells/ml • Or stage III or IV irrespective of CD4 count	• Stage II or III with CD4 ≤ 200 cells/ml • Or stage III with CD4 200 – 350 cells/ml • Or stage IV irrespective of CD4 count

Adapted from: World Health Organization, 2010a.

2.1.2. Current ARV drugs and ART regimens

There are currently 26 ARV drugs approved for clinical use, and these drugs are divided into six groups: nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), integrase strand transfer inhibitors (INSTIs), fusion inhibitors (FIs) and co-receptor inhibitors (CRIs) (Kumari & Singh, 2012; Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014). Tenofovir disoproxil fumarate (TDF) may be grouped separately as a nucleotide reverse transcriptase inhibitor (NtRTI) (Kumari & Singh, 2012).

The majority of HIV-positive patients receiving ART are prescribed a first-line ART regimen (World Health Organization, 2013a). The strategy for first-line ART in adults and adolescents is to prescribe a combination of three ARV drugs, as part of the triple-highly active antiretroviral therapy (HAART) (World Health Organization, 2006). This regimen usually consists of one NNRTI and two NRTIs, because it is effective in terms of cost, viral suppression and immunological regeneration, and can be prescribed as a fixed dosage, which can improve patient adherence (Martín *et al*., 2010; World Health Organization, 2010a; Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014, World Health Organization, 2014). Other recommended regimens can include two NRTIs and either one PI (especially for children less than three years of age) or one INSTI (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014).

Specialised regimens are prescribed for pregnant women, although discrepancies exist among leading countries and organisations. The DHHS recommends the use of nevirapine (NVP), as opposed to EFV, as the most appropriate NNRTI for pregnant and breastfeeding women (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014). Primate and human studies have reported growth and neuronal deformities in fetuses exposed to EFV in the first three months of pregnancy (Jeantils *et al.*, 2006; Bristol-Myers Squibb Pharma Company, 2013). This stance was supported by the 2006 WHO Guidelines; however, since 2010, the guidelines were altered in favour of EFV prescription for pregnant women (World Health Organization, 2006; 2010a). The motivation for this alteration was the increasing number of studies and review articles supporting the stability of EFV use and the increased risk for toxicity in pregnant women due to exposure to NVP (Ford *et al*., 2011; Ekouevi *et al*., 2011; Bera & Mia, 2012), as well as the WHO's aim of simplifying its ART guidelines. The British HIV Association (BHIVA) in 2012 and in the DHHS in 2013 also added the use of EFV as alternative NNRTI for HIV-infected pregnant women (Taylor *et al*., 2012; Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014).

The treatment prescribed for patients with TB-HIV and/or HBV-HIV co-infections should be carefully considered, due to ARV drug-drug interactions with the TB drug rifampicin, the common occurrence of immune reconstitution inflammatory syndrome (IRIS), the increase in progression of liver-disease in HBV-chronic patients and problems with drug toxicities. The WHO Guidelines recommend a standard regimen of one NNRTI and two NRTIs for TB-HIV- and/or HBV-HIV-infected patients. The NRTIs TDF and lamivudine (3TC) are preferred for HBV-HIV-infected patients, because of the antiviral activity against both HIV- and HBV-infections (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014; World Health Organization, 2014).

The recommended second-line ARV-regimen consists of one PI (enhanced with ritonavir (RTV)) and two NRTIs, providing that first-line therapy did not include a PI (World Health Organization, 2013b). PIs are preferentially used in second-line treatment because they differ substantially from NRTIs and NNRTIs in terms of structure, target viral receptor, and half-life, and are therefore suitable for combating viral resistance to first-line therapy (World Health Organization, 2006; Martín *et al*., 2010). It is important to select three active drugs for second-line therapy to aim for complete viral suppression, thus NRTIs are not re-prescribed (World Health Organization, 2013b). According to Hamers *et al.* (2012), tenofovir (TDF) and stavudine (d4T) should not be interchanged, as the same pattern of viral resistance is observed in patients taking both NRTIs, and thereby reducing the activity of the second-line regimen. New and second-generation ARVs,

e.g. etravirine, (an NNRTI), raltegravir (RAL) (an INSTI) and enfuvirtide (a FI) are examined in past, recent and on-going studies for use in second-line regimens (Lazzarin *et al.*, 2003; Gatell *et al.*, 2010; Katlama *et al.*, 2010). The second-generation NNRTIs etravirine and rilpivirine are already considered for second-line treatment in the DHHS guidelines (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014). However, in the Hamers *et al.* (2012) study, patients failing first-line NNRTI-based therapy showed a rapid increase in viral mutations, which suggests that second-line NNRTIs will have decreased efficiency when used as second-line treatment.

The introduction of rilpivirine, an alternative NNRTI to EFV and NVP, has recently received attention from the public and health sector (James *et al.*, 2012). Clinical trials report no significant difference between rilpivirine and EFV with regards to virological response and CD4 cell count, but rilpivirine resulted in fewer ADRs, with the associated reduction in treatment discontinuation. However, patients with higher viral load (VL) (> 10,000 copies/ml) presented an increase in the frequency of viral failure and, subsequently, an increase in ARV resistance (James *et al.*, 2012).

2.1.3. Evaluation of treatment outcome

ART theoretically suppresses all viral replication; however, treatment outcome differs among individuals and is dependent on a number of variables. Treatment outcome can be measured by CD4 cell count (immunological outcome), VL, clinical outcome and TDM (Woldemedhin & Wabe, 2012). The occurrence of adverse drug reactions (ADRs) in response to treatment is also an important aspect to consider when monitoring treatment outcome.

Patients will experience varying responses to treatment during the first few weeks that would not necessarily be indicative of treatment outcome, as the body is still adjusting to the treatment (World Health Organization, 2006; Lima *et al.*, 2012). These responses can manifest in different ways, for example sudden increases in VL and short-lived instances of ADRs (Havlir *et al.*, 2001). The onset of IRIS, the occurrence of symptoms of preceding diseases in patients because of an improving immune system in HIV patients, is also not an indicator of treatment failure (World Health Organization, 2006; Steele *et al.*, 2011). TB and cryptococcal disease are most common diseases that present as IRIS (World Health Organization, 2006). Accurate measurements of treatment outcomes are further complicated by the co-prescription of other medications for additional diseases. These drugs usually have their own set of adverse effects,

and can also interact with ARV medication to decrease the effectiveness of the ARV treatment (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014).

2.1.3.1. CD4 cell count

CD4 cell count is an indication of the immunological outcomes of a patient and can be used as a measurement of treatment response. According to the DHHS, it is also the best predictor for determining treatment response in a clinical environment (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014). A CD4-count of below-baseline or on-going counts below 100 cells/ml suggests immunological failure (World Health Organization, 2013b). However, a CD4 count increase of 50-150 cells/ml per year is considered to be a good immunological response (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014).

2.1.3.2. Viral load

Since 2010, HIV treatment studies using VL as measurement of treatment outcome have become more prevalent, and have shown that VL is a more sensitive measurement than CD4 count to identify treatment failure (Laurent *et al.*, 2011; Loutfy *et al.*, 2013). By 2013, VL was considered by the WHO to be the most important method of measuring treatment response (World Health Organization, 2013b).

Treatment response can range from successful viral suppression to virologic failure and viral resistance to drugs (Rodríguez-Nóvoa *et al.*, 2006). The cut-off value for virologic failure is still under discussion, and different values are being used by different studies. Currently, the WHO defines virologic failure as the occurrence of at least two sequential measures of VL above 1,000 copies/ml. In the 2010 guidelines, the cut-off value was 5,000 copies/ml (World Health Organization, 2010a). Alternatively, the DHHS differentiates between virologic failure (< 200 copies/ml) and persistent low-level viremia (< 1,000 copies/ml) (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014). One reason for the amended 1,000 copies/ml cut-off-value, is that viral transmission is significantly reduced in patients with VL below this value (Loutfy *et al.*, 2013). Some studies have reported that sporadic occurrences of VL levels between 50 and 1,000 copies/ml are not statistically associated with an increased risk of treatment failure, however, other studies claim that when these VL values occur consistently, they indicate treatment failure (Havlir *et al.*, 2001; Castelnuovo *et al.*, 2011; Lima *et al.*, 2012). Unfortunately, due to the high cost of scheduled VL measurements in patients (i.e. every three to six months), it is difficult to implement more frequent measurements in clinics.

2.1.3.3. Clinical outcome

The clinical results of a patient also indicate treatment response. Recent, new and on-going clinical complications, for example TB, pneumonia, fungal infections, and HBV-infections reported six months after ART initiation, can indicate viral failure (World Health Organization, 2013b). However, a problem with relying only on clinical outcome to identify treatment outcome, is that high viral loads can remain undetected, and, by the time patients are switched to another treatment regimen, viral resistance can already be present (World Health Organization, 2006).

2.1.3.4. Therapeutic Drug Monitoring

Therapeutic Drug Monitoring (TDM), the use of drug levels in the body to determine and regulate treatment outcome, is a possible additional approach to monitor ART (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014). TDM has been proven effective for the observation of plasma drug levels for PIs saquinavir (SQV), RTV, indinavir (IDV) and nelfinavir (NFV) in a research capacity (Schapiro *et al*., 1996; Durant *et al*., 2000). The use of TDM to trace the patient's reaction to treatment is especially useful for ARVs with a narrow therapeutic range and wide inter-patient variability, like NNRTIs EFV and NVP and the NRTI zidovudine (AZT) (Marzolini *et al*. 2001; Ståhle *et al*., 2004; Antonelli & Turriziani, 2012; Zanger & Schwab, 2013).

Plasma ARV levels are measured from blood samples collected from patients at regular intervals (every one to three months) using variations of high-performance liquid chromatography (HPLC) (Durant *et al*., 2000; Marzolini *et al*., 2001; Ståhle *et al*., 2004; Figueroa *et al*., 2010). These plasma ARV levels are then used, in conjunction with other evaluation methods, as an indication of patient response to treatment (Durant *et al*., 2000; Figueroa *et al*., 2010). ARV drug levels above the specific treatment range are an indication of the risk of occurrence of adverse effects (Haas *et al.*, 2004; Huang *et al.*, 2008; Gutiérrez *et al.*, 2013; Menezes *et al.*, 2013), while, sub-therapeutic drug levels may be an indication of decreased drug activity, treatment failure and the development of viral resistance (Marzolini *et al.*, 2001; Huang *et al.*, 2008).

Some organisations discourage the use of TDM in the clinical environment (Thompson *et al*., 2012). Measuring plasma ARV concentration can prove to be difficult, because of logistical barriers in obtaining a uniform sample in all patients. This is because of the varying pharmacokinetics of ARV drugs among patients, as well as varying time intervals between ingesting the drug and measuring plasma levels, differences in drug-drug interactions among

patients and differences in adherence among patients (Duval *et al.*, 2007). This approach is also subject to confounding factors, such as patient diet and metabolism.

Measuring ARV levels in hair, which represent drug levels in the body over a period of weeks or months, is an alternative method to consider (Huang *et al*., 2008). Duval *et al.* (2007) examined the effect of drug concentrations in plasma vs. hair levels for IDV, a protease inhibitor. After multivariate analysis, only hair samples were significantly associated with VL (Duval *et al.*, 2007). The practical advantages of hair samples are that they are easier and less invasive to collect and can be stored for longer periods of time. Huang *et al.* (2008) examined the available methods of extracting drug levels from hair samples and concluded that a technique combining liquid mass chromatography with gas spectrometry would most efficiently extract and measure EFV, lopinavir (LPV) and RTV levels in hair samples. A study by Gandhi *et al*. (2012) analysed the short-term effects of 230 polymorphisms on EFV levels by using plasma samples and the long-term effects by using patient hair samples. A total of 28 polymorphisms, including two important *CYP2B6* SNPs (*CYP2B6516G>T* and *CYP2B6983T>C*), had a reduced, but still significant, effect on EFV hair levels compared to EFV plasma levels (Gandhi *et al*., 2012). In a study published in 2013, Gandhi *et al*. used the same strategy of plasma levels representing recent exposure and hair levels representing long-term exposure, to show that EFV was transferred from mother to infant both through the placenta while in the uterus and through breastfeeding after birth, but LPV and RTV were transferred from mother to infant only through the placenta. Another study by Gandhi *et al*., (2014) compared two methods of measuring ARV treatment adherence in Kenyan patients: ARV hair levels and self-reported adherence. The study suggested that ARV levels in hair are a more accurate representation of treatment adherence than adherence reported by the patients themselves.

2.1.3.5. Adverse drug reactions

Drug toxicity (the advent of adverse reactions) is one of the biggest reasons for changing to an alternative ARV regimen (Kumarasamy *et al.*, 2006; Hart *et al.*, 2007; Woldemedhin & Wabe, 2012). Toxicity to ARV therapy is manifested in a wide range of ADRs, including hypersensitivity (caused by NRTIs and NNRTIs), central nervous system (CNS) side effects (caused by EFV), peripheral neuropathy (caused by d4T), nausea, anaemia, neutropenia (caused by AZT) back pain, double vision and, after long term exposure, hepatotoxicity (caused by NVP) and lipodystrophy (caused by d4T) (World Health Organization, 2006; Wyen *et al.*, 2011; Woldmedhin Wabe, 2012). These effects can occur within a week of treatment initiation, or at a

later stage, and can continue for several months regardless of treatment regimen alteration or discontinuation (World Health Organization, 2006; Wyen *et al*., 2011).

Studies disagree with regards to the most frequent cause of ART-related drug toxicity. Woldemedhin & Wabe (2012) studied patients who were prescribed a variety of treatment regimens, and identified peripheral neuropathy as the primary cause for altering treatment regimens. Alternatively, two other studies, one performed on Peruvian patients and one performed on North American patients, reported anaemia as the primary cause for altering treatment regimens (Bangsberg *et al.*, 2006; Nevin *et al.*, 2011).

Usually the severity of the ADRs determines whether a change of regimen is needed (Wyen *et al*., 2011). An effective and timeous intervention is vital, because the occurrence of any of the potential ADRs may lead to treatment non-adherence or discontinuation and increase the risk of drug resistance and/or virologic failure (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014). Drug toxicity is monitored through physician- and patient-reported symptoms during clinical assessments and laboratory analysis of blood samples (World Health Organization, 2013b). Toxicities are not always carefully observed and recorded, and therefore it is difficult to determine the absolute effect of these adverse reactions on patient treatment response and adherence (World Health Organization, 2006). The presence of the symptoms of co-occurring diseases (e.g. malaria or TB), as well as IRIS, also complicates the accurate annotation of ART-related adverse effects (World Health Organization, 2006).

A change in drug regimen owing to adverse effects may, however, lead to a greater risk of viral resistance (Nevin *et al.*, 2011). Viral resistance to ARVs results in poor treatment outcome, such as virologic failure and mortality (Nevin *et al.*, 2011). Standardised genotypic tests for viral resistance are available, testing for mutations in the genes encoding the viral reverse transcriptase (RT) and protease enzymes (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2014).

2.1.4. Antiretroviral therapy in a South African context

A report by Statistics South Africa (ZA) (2013) shows a prevalence of HIV-infected individuals of 10% in the South African population (Statistics South Africa, 2013). The latest data collected by the WHO shows that by the end of 2012, 2.15 million South African citizens were receiving ART, with an additional 2.7 million citizens eligible for ART (World Health Organization, 2013a). This amounts to a coverage of 83%, which is the third highest coverage of countries in Africa, following Botswana and Namibia (World Health Organization, 2013a).

Criteria for a South African HIV-positive patient to start ART are based on clinical as well as immunological assessments. National guidelines in 2010 were more stringent in terms of ART eligibility than the updated 2013 version (National Department of Health, 2010, 2013). Currently, all patients with a CD4 count of <350 cells/ml are eligible for ART; this correlates with current WHO standards. Priority is given to patients with a CD4-count below 200 cells/ml, co-infected with TB and who clinically present other WHO stage IV criteria (National Department of Health, 2013).

South African first- and second-line ART regimens are in accordance with WHO guidelines. The general first-line ARV-regimen prescribed for South African patients, is one NNRTI (EFV or NVP), along with two NNRTIs (TDF plus emtricitabine (FTC) or 3TC) (National Department of Health, 2013). Pregnant HIV-positive patients are currently also prescribed EFV, in contrast to 2010 South African guidelines that promoted the use of NVP in pregnant patients. Prior to 2010, South African patients were commonly prescribed d4T as part of the first-line regimen, however this drug was phased out because of severe adverse effects such as peripheral neuropathy, lipodystrophy, and pancreatitis (World Health Organization, 2010a). A second-line regimen consists of a RTV-boosted PI-based regimen (including LPV or atazanavir as the PI), and patients with a VL above 1,000 copies/ml are eligible for this regimen (National Department of Health, 2013). A third-line ART option exists that includes new ARV drugs such as darunavir and RAL.

South Africa first implemented a national ART-plan in 2004 and even though this was an important achievement, there are still some problems in the clinical environment and also in the patients' living environment that need to be addressed to access the full potential of ART (World Health Organization, 2010b). The limited number of doctors and nurses available to treat and monitor the vast number of HIV-infected patients in South Africa is of primary concern (Fairall *et al.*, 2012), thereby limiting the potential improvement in the quality of patient support, counselling and care that are required (Harries *et al.*, 2001). In addition, the regulation of access to ARVs for eligible and non-eligible patients is compromised, potentially leading to death for eligible patients that do not start ART treatment early enough or drug resistance in treated patients who are not yet eligible (Harries *et al.*, 2001; Fairall *et al.*, 2012). A strategy has been proposed and tested in some South African clinics in which the workload is shared between doctors and nurses, with nurses being assigned to prescribe ARV medication, while doctors deal primarily with patients displaying severe disease symptoms (Fairall *et al.*, 2012). This strategy proved to be successful, and is currently being implemented in other South African clinics.

Another challenge specific to sub-Saharan African countries, including South Africa, is the regulation of eligibility for second-line treatment regimens and controlling the emergence of viral drug resistance. Although South African Guidelines state that VL should be used as a measurement of eligibility for second-line ART, this is not yet fully implemented in South Africa and VL is not yet efficiently monitored, according to WHO standards (Harries *et al*., 2010; World Health Organization, 2013b). Some clinical studies in sub-Saharan Africa reported that 24% of HIV patients have a median VL of 400-500 copies/ml after six to 36 months on ART (Wester *et al.*, 2005; Ferradini *et al.*, 2006; Boulle, 2008). However, these findings tend to include the presence of viral blips (short occurrences of high VL), which are not necessarily indicative of viral failure.

Patients' living environments can also complicate treatment. The vast majority of South African HIV-positive patients live in poverty, which may result in a lack of nutrition, clean drinking water, sanitation and shelter (United Nations, 1995; Evans *et al*., 2012). Only malnutrition, however, has been associated with an increase in the severity of HIV disease (Macallan, 1999; Bates *et al.*, 2004; Evans *et al.*, 2012). Evans *et al.* (2012) recruited patients from Johannesburg in South Africa that were prescribed an ARV regimen as suggested in national guidelines and reported that a low body mass index (BMI), as a result of malnutrition, was significantly associated with increased VL, an increased chance of developing TB, increased onset of oropharyngeal candidiasis (OC), an HIV-disease-related symptom, and an increased risk of death.

2.1.5. Efavirenz

EFV is an ARV drug used in the treatment of HIV-1 positive patients. It is an NNRTI, the group of ARVs that inhibit the activity of the RT enzyme of HIV-1, thereby hindering the conversion of viral RNA into DNA for integration into the host DNA (Ward *et al.*, 2003; WHO Collaborating Centre, 2011). Reverse transcriptase (RT) gene mutations, most commonly K103N in the HIV-1 strain, are responsible for EFV drug resistance, in that it reduces viral susceptibility to the drug (Adkins & Noble, 1998; Mollan *et al.*, 2012; Thiam *et al.*, 2013).

2.1.5.1. The pharmacokinetics of EFV and the factors influencing EFV metabolism

EFV ($C_{14}H_9ClF_3NO_2$) has a molecular mass of 315.68 and is ingested orally by patients in tablet form (600 mg), or in capsule form (50 mg or 200 mg) (Bristol-Myers Squibb, 2013). EFV, like most types of xenobiotic substances, is lipophilic and can enter intestinal cells through the cell membrane (Lindenberg *et al*., 2004; Bristol-Meyers Squibb, 2013). In plasma, EFV binds

predominantly to albumin (Bristol-Myers Squibb, 2013), with a very small percentage (0.4 to 1.5%) unbound to proteins (Almond *et al*., 2005; Avery *et al*., 2013). In contrast to PIs, EFV's small molecular mass enables the drug to cross the blood-brain barrier and enter the cerebrospinal fluid (CSF), with a plasma-CSF-gradient of 150 (Best *et al.*, 2011). Although the EFV-concentration in the CSF is much lower than in plasma, it is still higher than the EFV IC_{50}-value (26 times higher), and is therefore sufficient for treatment success (Tashima *et al*., 1999; Best *et al*., 2011).

EFV is primarily metabolised in the liver by CYP2B6 with assistance from CYP3A4, CYP3A5, CYP2A6, CYP1A2 and UGT2B7 to its inactive and conjugated forms, 8-hydroxyEFV, 7-hydroxyEFV, 7,8-dihydroxyEFV and EFV-*N*-glucuronide (Figure 1) (Mutlib *et al*., 1999; Bristol-Myers Squibb, 2013). CYP2B6 is also responsible for the secondary metabolism of EFV to 8,14-dihydroxyefavirenz (Ward *et al.*, 2003). Hepatocytes make use of these oxidative pathways to metabolise EFV into hydrophilic compounds, which facilitates excretion via urine (Lee, 2003).

EFV has a plasma half-life of approximately 40 to 76 hours and has a plasma therapeutic range of 1 to 4 μg/ml (Marzolini *et al.*, 2001; Castillo *et al.*, 2002; Haas *et al.*, 2004; Bristol-Myers Squibb, 2013). The long half-life of EFV provides sufficient time for the drug be effective, but the narrow therapeutic range of EFV reduces the chance of good treatment outcome. Low, or sub-therapeutic EFV plasma levels (< 1 μg/ml) are associated with a higher risk of treatment failure and drug resistance and may be due to an increased EFV clearance rate in certain individuals (Ward *et al.*, 2003; Tsuchiya *et al.*, 2004; Rodríguez-Nóvoa *et al.*, 2006). High, or super-therapeutic EFV plasma levels (> 4 μg/ml) may be the result of a slow EFV clearance rate and increases the probability of adverse effects, especially those that impact the CNS, including headaches, dizziness, insomnia and fatigue (Marzolini *et al*., 2001; Haas *et al*., 2004; World Health Organization, 2014). Additional adverse effects of EFV include hepatotoxicity, rash, hypersensitivity, dyslipidaemia, nausea, heartburn and diarrhoea (Adkins & Noble, 1998; Shubber *et al*., 2013; World Health Organization, 2013c). There is substantial inter-individual variability in EFV levels, which results in different toxicity patterns, further complicating the research aim of optimising EFV treatment (Haas *et al.*, 2004).

Figure 1 illustrates the various factors that can influence EFV metabolism. HIV-infected patients taking EFV is often prescribed additional medication for other diseases, therefore the effects of drug-drug interactions must be taken into account. The enzymes responsible for EFV metabolism are also involved in the metabolism of other drugs, including IDV and SQV (PIs), AZT (an NRTI) rifampicin (anti-TB drug), clarithromycin (antibiotic), pethidine (pain relieving drug) and

artemisin (antimalarial drug) (Ramírez *et al*., 2004; Asimus & Ashton, 2009; Belanger *et al*., 2009; Gengiah *et al*., 2012; Hasegawa *et al*., 2012; Huang *et al*., 2012). In some cases, the co-administration of these drugs results in impaired metabolism of EFV or the other drugs involved (Belanger *et al*., 2009; Gengiah *et al*., 2012; Hasegawa *et al*., 2012; Huang *et al*., 2012).

Several xenobiotics have also been reported to inhibit the activity of EFV-metabolising enzymes either through direct binding or through indirect activation of nuclear receptors (NRs) and other transcription factors (TFs). These xenobiotics include clopidogrel (antiplatelet drug), methadone (pain relieving drug) and tamoxifen (hormone therapy drug) (Richter *et al*., 2004; Amunugama *et al*., 2012; Sridar *et al*., 2012).

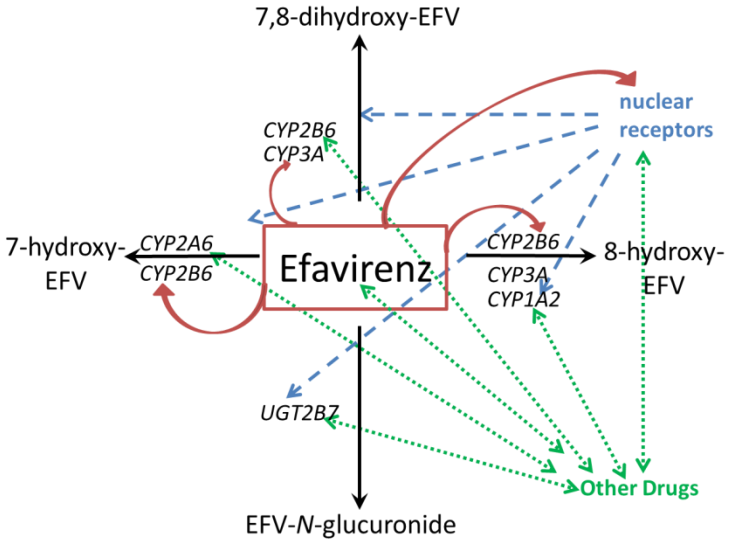


Figure 1: Diagram representing the metabolism of EFV and the factors that influence the rate of EFV metabolism.
The red arrows indicate the influence of EFV on the regulation of its metabolising enzymes. The green arrows indicate the influence of other drugs on EFV metabolism and *vice versa*. The blue arrows indicate the influence of nuclear receptors and other transcription factors on EFV metabolism through the binding of EFV itself and other drugs. Adapted from Di Iulio *et al*., 2009, with publisher's permission.

EFV also influences the activity and transcription of its metabolising enzymes, *CYP2B6* and CYP3A4, through an indirect mechanism by binding to NRs (Hariparsad *et al*., 2004; Habtewold *et al*., 2012). Therefore, EFV induces its own metabolism. The above-mentioned factors show that further studies are required in order to completely understand all the complexities involved in EFV metabolism.

2.1.5.2. The pharmacogenomics of EFV

Genetic variants in all genes playing important roles in EFV metabolism (Figure 1), as well as the genes encoding potential transporters for EFV, have been implicated in pharmacogenomic studies on EFV.

Genetic variants in the gene encoding *CYP2B6*, the main metaboliser of EFV, influence EFV levels in the body, treatment outcome and the occurrence of ADRs (Haas *et al*., 2004; Motsinger *et al*., 2006; Habtewold *et al*., 2011; Gandhi *et al*., 2012; Arab-Alameddine *et al*., 2013; Martín *et al*., 2013). Haas *et al.* (2004) assessed the interaction between genetic variation, EFV plasma levels and CNS-specific ADRs in a multicultural cohort of African American, Latin-American and Caucasian HIV-positive patients receiving EFV treatment and identified an association between the T-allele of *CYP2B6* SNP 516G>T (part of the *CYP2B6*6* allele) and higher EFV plasma exposure over a period of six months. This association was replicated in studies that utilised different methods of measuring EFV levels, including long-term EFV plasma exposure, EFV plasma clearance after a single dose, EFV levels in hair, and intracellular EFV levels in peripheral blood mononuclear cells (PBMCs) (Haas *et al.*, 2005; Wang *et al.*, 2006; Haas *et al*., 2009; Habtewold *et al.*, 2011; Gandhi *et al.*, 2012). An association was also found between the C-allele of another *CYP2B6* SNP, 983T>C, (part of the *CYP2B6*16* and **18* alleles) that have only been identified in individuals of African or admixed-African ethnicity, and elevated EFV plasma levels (Wyen *et al.*, 2008; Gandhi *et al.*, 2012; Maimbo *et al.*, 2012; Mutwa *et al.*, 2012; Johnson *et al.*, 2013; Swart *et al.*, 2013).

Arab-Alameddine *et al.* (2009) developed a pharmacokinetic model for EFV clearance that identified two *CYP3A4* SNPs, namely rs274057 (4713G>A) and rs4646437 (21726C>T), associated with decreased EFV clearance in combination with *CYP2B6*6*. *CYP3A4* rs4646437 was also associated with an increase in the occurrence of EFV discontinuation as a result of treatment failure or ADRs (Lubomirov *et al.*, 2011).

ABCB1 encodes the transporter protein p-glycoprotein (P-gp), and although studies report that P-gp does not play a direct role in the transportation of EFV (Burhenne *et al.*, 2010), the influence of variants in *ABCB1* on EFV plasma levels warrants further investigation. Several studies identified an association between an *ABCB1*-SNP, 3435C>T, and a decreased chance of virologic failure, viral resistance, drug toxicity and ADRs in patients receiving EFV (Fellay *et al.*, 2002; Haas *et al.*, 2005; Motsinger *et al.*, 2006; Ritchie *et al.*, 2006; Yimer *et al.*, 2011).

Other genes that have also indicated correlations among genomic variation, EFV levels and treatment outcome are *CYP1A2*, *CYP2A6*, *UGT2B7*, and two NR genes, *NR1I3* and *NR1I2* (Kwara *et al*., 2009; Habtewold *et al*., 2011; Wyen *et al*., 2011; Swart *et al*., 2012; Cortes *et al*., 2013; Martín *et al*., 2013).

2.2. Nuclear hormone receptors

2.2.1. Structure and organisation of the nuclear receptor superfamily

Nuclear hormone receptors (NRs) form a family of TFs that regulate the transcription of genes involved in the metabolism of xenobiotic compounds as well as various endogenous compounds (Lamba *et al*., 2008; Ihunnah *et al*., 2011). NRs are also involved in vital cellular processes such as cell growth and differentiation, chromatin condensation, gene silencing, physiological homeostasis and metamorphosis (Mangelsdorf *et al*., 1995; Mckenna & O'Malley, 2002; Bertrand *et al*., 2004).

NRs are only found in metazoans and are phylogenetically separated into six subfamilies (NR1 to NR6) (Giguère, 1999; Nuclear Receptors Nomenclature Committee, 1999). The number of NRs differs greatly among species, although trends can be observed in closely related species. For example, humans and mice have 48 and 49 NRs respectively, whereas the mosquito and the fruit fly have 21 NRs each, and nematodes *Caenorhabditis elegans* and *C. briggsae* both have more than 250 NRs (Bertrand *et al*., 2004). The diversity in total number of NRs among species can be explained by gene duplication periods as well as a loss of irrelevant genes with the passing of time (Maglich *et al*., 2001; Bertrand *et al*, 2004).

Maglich *et al.* (2001) investigated the phylogenetic differences among nuclear receptor genes of two invertebrate and one vertebrate species using genome information obtained from the nematode (*C. elegans*), the fruit fly (*Drosophila melanogaster*) and humans. They reported several NRs present in humans that are absent in invertebrates, including thyroid hormone receptors, androgen receptors and steroid receptors. Additionally, 24 NRs of *C. elegans* and three NRs of *D. melanogaster* could not be classified according to any of the existing six subfamilies. The NR3 class contains no representatives of *C. elegans* and only one of the *D. melanogaster*, whereas the human NRs are represented in all six NR-subfamilies.

NRs can also be separated into two groups based on their method of identification and ligand binding affinity. NRs that have an affinity for specific physiological ligands and were identified through protein and hormone binding studies are called ligand-dependent nuclear receptors. A

large number of NRs were identified through hybridisation and cloning studies using probes from the highly conserved DNA binding domain of other NRs. These NRs are termed orphan nuclear receptors and are named as such either because they do not require binding ligands for activation (constitutive expression) or because their specific ligands have not yet been determined (Giguère, 1999). Orphan nuclear receptors account for 60% of all mammalian NRs and include the hepatocyte nuclear factor-4α (HNF4α), CAR and PXR (Mangelsdorf *et al*., 1995; Giguère, 1999; Chen *et al*., 2012).

The important structural features of nuclear receptors are the DNA binding domain (DBD), the ligand binding domain (LBD), and two transactivation domains (AF-1 and AF-2) (Mangelsdorf & Evans, 1995; Wang & LeCluyse, 2003; Chen *et al.*, 2012). The DBD is generally a highly conserved module among different species and plays a role in identifying and binding to xenobiotic response elements (XREs) or hormone response elements (HREs) on target genes (Mangelsdorf & Evans, 1995; Wang & LeCluyse, 2003; Lamba *et al*., 2008). The LBD is less conserved than the DBD and is involved in the binding of lipophilic ligands to the NR, which regulates the activity of NRs (Maglich *et al.*, 2001). The LBD also plays a role in the interaction with co-repressors or co-activators, dimerisation and nuclear localisation (Wang & LeCluyse, 2003; Lamba *et al*., 2008). The binding of a ligand induces a conformational change within the LBD, aided by AF-2, allowing the affinity of NRs for co-activators that aid in transcriptional regulation of target genes (Mangelsdorf *et al.*, 1995; Giguère, 1999). Some NRs, mostly from the NR1-subfamily, form heterodimers with the retinoid X receptor (RXR) when binding to XREs on their target genes (Mangelsdorf & Evans, 1995; Wang & LeCluyse, 2003).

In terms of species conservation of the DBD and LBD in NRs, CAR and PXR are among the exceptions. The DBD and LBD of PXR and, to a lesser extent, CAR, are less conserved among species than other NRs, on a protein as well as a genetic level. This explains the ability of human CAR and PXR to bind to a wider range of ligands than their homologs in other species (Reschly & Krasowski, 2006).

2.2.2. The role of CAR and PXR as transcriptional regulators in xenobiotic metabolism

CAR, encoded by *NR1I3*, and PXR, encoded by *NR1I2*, are the focus of this study, and overlap extensively in ligand specificity and target genes. CAR and PXR are involved in the transcriptional regulation of genes encoding for Phase I and II drug metabolising enzymes, genes involved in drug transport, steroid metabolism and bile acid and bilirubin detoxification, among

others (Xie *et al*., 2001; Handschin & Meyer, 2003; Lamba *et al*., 2008; Ihunnah *et al*., 2011). CAR and PXR are both expressed mainly in the liver (Wang & LeCluyse, 2003), which correlates with the expression of the key genes responsible for xenobiotic metabolism, including *CYP2B6*, *CYP3A4*, *CYP3A5* and *CYP2A6* (Chang *et al*., 2003; Lamba *et al*., 2003).

The enzymes in the CYP1, CYP2 and CYP3-subfamilies are responsible for the phase I metabolism of approximately 70 to 80% of xenobiotic substances metabolised in the liver (Figure 2) (Zanger *et al*, 2008). Maglich *et al.* (2002) reported that CAR and PXR enhance the expression of *CYP1A2* through ligand activation (Maglich *et al*., 2002). However, recent studies only observe the ability of CAR to influence *CYP1A2* expression (Yoshinari *et al*. 2010; Tojima *et al*., 2012) *In vitro* reporter assays by Yoshinari *et al*. (2010) showed that CAR plays a role in the activation of *CYP1A2* expression by binding to a XRE approximately 300 bp upstream from the start of the *CYP1A2* gene. PXR, however, is an important contributor to the transcriptional regulation of CYP3A4 and CYP3A5 enzymes, which play a role in the metabolism of 37% of xenobiotic substances (Zanger *et al*., 2008). Burk *et al.* (2004) observed that PXR induces *CYP3A4* and *CYP3A5* transcription in liver tissue by means of activation by the anti-TB drug, rifampicin (Burk *et al*., 2004). The study also reported that CAR, in cooperation with HNF4α, plays a role in the basal expression of *CYP3A* genes.

CYP2B6 is an important enzyme with respect to the metabolism of medical and environmental chemicals. Tissue culture studies demonstrate the phenobarbital-mediated activation of *CYP2B6* through a CAR-RXR-heterodimer binding to a 5'upstream region of the gene, called the phenobarbital-responsive enhancer module (PBREM) (Sueyoshi *et al*., 1999). Wang *et al.* (2003) also reported the influence of PXR on *CYP2B6* regulation through the discovery of yet another response element 5'upstream of the PBREM element that is capable of binding to both CAR and PXR. Using cultured liver cells, both CAR and PXR were shown to induce *CYP2B6* expression through binding to both response elements, PBREM and the novel response element.

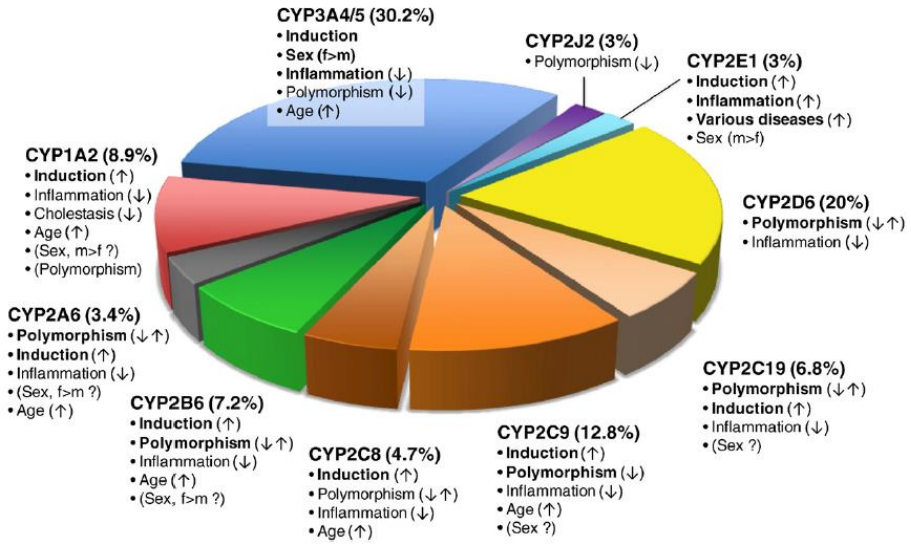


Figure 2: Representation of the Cytochrome P450 enzymes involved in drug metabolism and the factors that influence their activity
The percentage values indicate the proportion of drugs that are metabolised by the specific enzyme. ↑ indicates increased enzyme activity. ↓ indicates decreased enzyme activity. Taken from Zanger & Schwab, 2013, with publisher's permission.

Additional CYP-enzymes that are regulated by CAR and PXR include CYP2A6, CYP2C9, CYP2C19 and CYP4F12 (Chen *et al.*, 2005; Tolson *et al.*, 2009). Other phase I enzymes that are regulated by PXR include aldo-keto reductases (AKRs), enzymes that break down toxic aldehydes and ketones, and carboxylesterases (CESs), enzymes that break down esters, amides and thioesters (Rosenfeld *et al.*, 2003).

UGT-enzymes are involved in the glucuronidation of xenobiotics and endobiotics as part of phase II metabolism. A member of this family of enzymes, UGT1A1, is responsible for the metabolism and clearance of bilirubin in the body. Xie *et al.* (2002), using transgenic mice, observed that PXR activated by rifampicin increased the transcription and expression of *UGT1A1*. Sugatani *et al.* (2005), using cultured oncogenic liver cells, also observed that CAR and PXR induce *UGT1A1* transcription through dexamethasone activation and co-activation through the glucocorticoid receptor (GR). CAR and PXR also regulate other UGTs such as UGT1A4 and UGT1A9 (Ji *et al.*, 2012).

Other Phase II enzymes that CAR and PXR regulate, include sulfotransferases (SULTs), enzymes that are responsible for the conjugation of substances through the addition of a sulfonyl group, and glutathione S-transferases (GSTs), enzymes that catalyses the addition of glutathione substances (Xie *et al.*, 2002; Sugatani *et al.*, 2005; Duret *et al.*, 2006; Alnouti & Klaassen, 2008; Sueyoshi *et al.*, 2011).

Transporter proteins play a significant role in controlling homeostasis for both xenobiotic and endobiotic compounds in a cell. Geick *et al.* (2001) used a cultured colon cancer cell line to show that *ABCB1* expression is induced by rifampicin through the direct binding of PXR to a response element in the *ABCB1* upstream region. CAR is also capable of binding to *ABCB1* and influencing its expression in colon cancer cell lines (Burk *et al*., 2005).

Some of the ligands that are recognised by CAR and PXR include steroids (e.g. pregnane), catabolites (e.g. bilirubin), herbal xenobiotics (e.g. hyperforin) and drugs used in various clinical treatment regimens (e.g. rifampicin) (Di Masi *et al.*, 2009). Various ARVs are also ligands of CAR and PXR, including the PIs, IDV and RTV, and NNRTIs, NVP and EFV (Sinz *et al.*, 2006; Faucette *et al.*, 2007). However, the affinity of PXR for EFV is weak, as observed in the analysis of PXR-activity using reporter assays. PXR-activity in the presence of EFV was only 30% when measured as a percentage of total PXR-activity in the presence of rifampicin (Faucette *et al.*, 2007). CAR has a strong affinity for EFV and is up-regulated by EFV by as much as ten times the basal expression ($P < 0.05$) (Faucette *et al.*, 2007).

Both *NR1I3* and *NR1I2* undergo alternative splicing to form different isoforms of CAR and PXR, resulting in different structural and functional characteristics, which might help to explain why CAR and PXR can recognise such a variety of ligands and target genes. Di Masi *et al.* (2009) suggested that SNPs in the critical splice areas of these genes could affect protein activity, and therefore also affect the biological pathways containing CAR and PXR.

2.2.3. CAR encoded by *NR1I3*

NR1I3 is 8,600 base pairs (bp) in length and is located on chromosome 1, region q23.3 (Lamba *et al*., 2004; Ensembl Genome Browser). The gene consists of nine exons, the first of which forms part of the 5'-untranslated region (UTR). The remaining eight exons can undergo alternative splicing to form the wild type CAR and 26 different isoforms (Arnold *et al*., 2004; Lamba *et al*., 2004). The wild-type CAR-protein, CAR.1, is translated from a mRNA molecule, 1,153 nucleotides (nt) in length, that contains the transcription product of all nine exons (Lamba *et al.*, 2004). CAR.1 consists of 348 amino acids (aa) and has the highest ligand binding affinity of all the CAR isoforms (Lamba *et al.*, 2004). There are some CAR-isoforms that incorporate premature stop-codons, and are therefore translated to form non-functional proteins that are rapidly degraded (Lamba *et al.*, 2004; Di Masi *et al.*, 2009). CAR is constitutively expressed, because it allows nuclear co-activators to bind without the necessity for concurrent ligand binding

(Xiao *et al*., 2010), however, the activity of CAR can be up- and down-regulated through the binding of specific ligands and TFs (Pascussi *et al*., 2003; Assenat *et al*., 2004).

2.2.3.1. Population distribution and function of genetic variation in NR1I3

Until recently, only three main studies provided data for genetic variation in the gene (Ikeda *et al*., 2003, 2005; Thompson *et al*., 2005). Ikeda *et al.* (2003) reported 26 novel SNPs in a Japanese cohort of 253 subjects. Most SNPs were intronic and upstream variants, however, three SNPs were located in exons 4, 5, and 9 respectively; one was non-synonymous (V133G) and two were synonymous (P180P and Q344Q). However, no functional studies were performed. A follow-up study by Ikeda *et al.* (2005), involving functional studies as well as sequencing of *NR1I3* in 334 Japanese cancer patients, reported three novel exonic SNPs (H246A, L308P and N323S) with low allele frequencies (0.5%, 0.3% and 0.3% respectively). This study identified an association between the H246A and L308P variants and a decrease in *CYP3A4*-activation. Because this reduction was not due to reduced CAR expression, it was suggested that the two variants potentially cause a functional alteration of the protein.

Thompson *et al.* (2005) sequenced the *NR1I3* gene in 24 Caucasian, 23 African American and 23 Chinese Han individuals. The majority of variants that were identified occurred at a low frequency and were unique to each population, with the minor allele frequency (MAF) greater than 0.2 for only five of the 21 SNPs, and only five of the 21 SNPs occurred in all three ethnic groups. Four novel exonic variants were identified, including a non-synonymous SNP, R97W. *In silico* analysis reported that R97W disrupts a predicted exonic splicing enhancer (ESE).

The results published by two large-scale population genetic projects, the HapMap Project (The International HapMap 3 Consortium, 2010) and the 1000 Genomes Project (The 1000 Genomes Consortium, 2012), greatly increased the available variant data for most genes in the human genome, including *NR1I3*. To date, the Ensembl database, containing SNP data from both the HapMap Project and the 1000 Genomes Project, lists approximately 40 non-synonymous and 18 synonymous SNPs in *NR1I3*, a frequency of one variant per 18 nucleotides in the coding region. When compared to the characteristics of an average human gene, the SNP frequency of *NR1I3* is much higher. The *E2F2* gene is representative of the average characteristics of 213 genes investigated in the Environmental Genome Project (Livingston *et al.*, 2004). This gene has a SNP frequency of one variant per 24 nucleotides in the coding region, which is 33% lower than the *NR1I3* SNP frequency (Ensembl Genomes Browser). These facts indicate that sequence conservation is lower in the coding region of human *NR1I3* than previously thought (Ikeda *et al*.,

2003; Thompson *et al*., 2005; Reschly & Krasowski, 2006). However, when investigating genetic conservation, a distinction should be made between common and rare variants, and differences among ethnic groups should be taken into account.

2.2.3.2. The role of NR1I3 variants in EFV metabolism and treatment outcome

CAR plays a role in the transcriptional regulation of most enzymes involved in EFV metabolism; it is therefore hypothesised that some genetic variation in *NR1I3* may influence the rate of EFV metabolism (Wang *et al*., 2003; Burk *et al*., 2004; Yoshinari *et al*. 2010). This could influence EFV levels *in vivo* and, subsequently, could have downstream implications, such as altered clinical outcome. Two *NR1I3* SNPs, rs2307424 and rs3003596, have been implicated in altered EFV levels or treatment outcome in previous studies. Table 2 lists the MAFs for these two SNPs in Caucasian, Asian, African and African American patients, taken from 1000 Genome data. Both SNPs occur frequently in all four ethnic groups, however, rs2307424 occurs at a lower frequency in African populations (MAF = 0.108). In addition, the minor allele for the Asian population (C) differs from the minor alleles in the other ethnic groups (T).

Wyen *et al.* (2011) investigated associations between *CYP2B6* and *NR1I3* polymorphisms and early EFV treatment default and discontinuation (< three months). The study used a cohort of 373 patients, the majority of which were of Caucasian ethnicity and the remainder were of African ethnicity. Out of the total cohort, 131 patients discontinued EFV-based therapy within the first three months (35%), for various reported reasons, including CNS-specific ADRs (55%). Other reasons for discontinuation include viral failure and hepatotoxicity, but a large percentage of patients (42%) defaulted for unknown reasons. After logistic regression analysis correcting for confounding factors, *CYP2B6* 516G>T and *NR1I3* rs2307424 were significantly associated with early EFV default (P = 0.006 and 0.007 respectively). Results also indicated that patients with composite genotypes for both SNPs were at a significantly higher risk of early default and discontinuation (P = 0.02). The confounding effects that were significantly associated with early EFV default are ethnicity and smoking habits (P < 0.05 in both cases). This is the only EFV-based study that analysed direct genotype-to-treatment outcome correlation, without using EFV levels as a transitional measurement.

Table 2: Population and functional details of *NR1I3* SNPs associated with EFV levels and treatment outcome in populations of Caucasian (Cau), Asian (Asn), African (Afr) and African American (AA) ethnicities.

Location	rs number	Genomic identifier	MAF				Functional Relevance	Reference
			Cau (CEU)	Asn (CHB)	Afr (YRI)	AA (ASW)		
Intron 4	rs3003596	8784T>C	0.453 (T)	0.420 (C)	0.386 (T)	0.492 (T)	Risk of CNS-ADRs, EFV plasma levels	Swart *et al.*, 2012; Mukonzo *et al.*, 2013
Exon 5	rs2307424	10396C>T	0.300 (T)	0.485 (C)	0.108 (T)	0.213 (T)	EFV discontinuation, EFV plasma levels	Wyen *et al.*, 2011; Swart *et al.*, 2012; Cortes *et al.*, 2013

CEU, CHB, YRI and ASW, represent four respective ethnic groups that form part of the 1000 Genomes Project. CEU, Residents in Utah, USA, of Northern and Western European ancestry. CHB, Han Chinese in Bejing, China. YRI, Yoruba in Ibadan, Nigeria. ASW, Americans of African Ancestry in South West, USA.

The rs2307424 SNP was also associated with EFV plasma levels in a Latin American cohort of 208 HIV-positive individuals (Cortes *et al.*, 2013). Patients with the rs2307424 T/T genotype had significantly decreased EFV plasma levels in comparison with the C/C ($P < 0.05$) and C/T ($P < 0.01$) genotype.

Swart *et al*. (2012), studying South African Bantu HIV-positive individuals (n = 464), also reported that rs2307424 is associated with increased risk of EFV discontinuation. In this study, 32% of the patients with the rs2307424 C/T genotype experienced discontinuation of EFV treatment within one year, as opposed to 21% of patients with the wild type (C/C) genotype. However rs2307424 was not significantly associated with EFV plasma levels in this study.

A SNP located in intron 4 of *NR1I3*, rs3003596 (8784T>C), was associated with EFV plasma levels in the Swart *et al.* (2012) study. Patients with the C/C and C/T genotypes had significantly lower EFV plasma levels than patients with the T/T genotype ($P = 0.015$ and 0.010 respectively). This significant association was not replicated in Mukonzo *et al.* (2013), a study of 197 patients living in Uganda. However, a trend was found between the rs3003596T/T and C/T genotypes and an increased risk of developing CNS-specific ADRs ($P = 0.06$). This corroborates previous studies that suggest that higher EFV levels lead to an increased occurrence of ADRs.

These results suggest that EFV plasma levels are not always a direct indication of treatment outcome and that the addition of other measurements of treatment outcome is needed in ARV pharmacokinetic studies. It is also suggested that the assessment of the influence of confounding factors on EFV levels is important, such as age, ethnicity and treatment adherence.

2.2.4. PXR encoded by *NR1I2*

NR1I2 is 34,500 bp in length and is located on human chromosome 3, region q13.33 (Hustert *et al*., 2001; Ensembl Genomes Browser). The gene contains ten exons, which are alternatively spliced to form the wild type PXR protein and nine other isoforms of PXR (Hustert *et al*., 2001). The wild-type protein, hPXR, contains translated material from all ten *NR1I2*-exons and is 434 aa in length (Lehmann *et al.*, 1998; Hustert *et al.*, 2001). The DBD consists of exons 3 and 4, while the LBD spans over exons 5 to 10 (Di Masi *et al.*, 2009). The translation of an important isoform of PXR, hPXR.2, differs from the wild-type PXR in that the 37 aa in the ligand binding domain is lost, resulting in decreased PXR activity (Hustert *et al.*, 2001).

Even though it is not constitutively expressed, and various physiological compounds have been identified as ligands, PXR is still classified as an orphan nuclear receptor (Kliewer *et al.*, 1998; Pondugula & Mani, 2013). Northern blot and reverse transcriptase-PCR (RT-PCR) assays report that PXR is expressed predominantly in the liver, and at a lower level in the colon and small intestine (Bertilsson *et al.*, 1998; Lehmann *et al.*, 1998). This differs from the expression patterns of PXR in mice (mPXR), where PXR is also expressed in stomach and kidney tissues, albeit at a much lower level (Kliewer *et al.*, 1998).

The ligand characteristics also differ between human and mouse or rat PXR homologs. For example, the affinity for rifampicin is much higher in human PXR than mouse PXR (Jones *et al.*, 2000; LeCluyse, 2001). This complicates the use of mouse models in PXR expression and association studies. However, Hasegawa *et al.* (2012) was able to circumvent this problem by using humanised mice to investigate the influence of PXR in the expression of CYP3A and CYP2C.

2.2.4.1. Population distribution and function of genetic variation in NR1I2

NR1I2 is less conserved among humans than *NR1I3*. The Ensembl-database, including data from the HapMap Project and the 1000 Genomes Project, currently lists approximately 65 non-synonymous and 31 synonymous SNPs in *NR1I2*, a rate of one variant per 15 nucleotides in the coding region. Table 3 compares *NR1I2* SNPs among four different ethnic groups from the 1000 Genomes Project and literature.

Table 3: MAF of *NR1I2* SNPs in populations of Caucasian (Cau), Asian (Asn), African (Afr) and African American (AA) ethnicities.

Location	rs number	Genomic identifier	Amino acid change, allele	MAF			
				Cau (CEU)	Asn (CHB)	Afr (YRI)	AA (ASW)
5'-upstream	rs1523130	5177T>C	n/a	0.359 (T)	0.268 (T)	0.040 (C)	0.139 (C)
5'-upstream	rs3814055	5705C>T	n/a	0.341	0.268	0.301	0.238
5'-UTR	rs1523127	6709C>A	n/a	0.341 (C)	0.268 (C)	0.097 (A)	0.230 (A)
Intron 1	rs13085558	18885T>C	n/a	0.188	0.178	0.023	0.066
Intron 1	rs2472679	30120A>T	n/a	0.188	0.077	0.489	0.418
Intron 1	rs2472677	24087C>T	n/a	0.359	0.371	0.352	0.361
Exon 2	rs12721613	31846C>T	P27S, *PXR**2	0.000	0.000	0.188	0.107
Exon 2	rs12721607	31873G>A	G36R, *PXR**3	0.000	0.000	0.000	0.016
Exon 4	rs12721608[1]	36089G>A	R122Q, *PXR*4*	0.010	n/a	n/a	0.000
Exon 4	rs72551372[2]	36142G>A	V140M, *PXR**10	0.001	n/a	0.002	n/a
Exon 4	rs72551374[3]	36212A>G	D163G, *PXR**11	0.000	n/a	0.014	n/a
Exon 8	rs35761343	40296G>A	A370T	0.000	0.000	0.023	0.016

[1]2001 Zhang *et al.* [2]Svärd *et al.*, 2010. [3]Hustert *et al.*, 2001. MAFs for all other SNPs obtained from the 1000 Genomes Database. CEU, CHB, YRI and ASW, represent four respective ethnic groups that form part of the 1000 Genomes Project. CEU, Residents in Utah, USA, of Northern and Western European ancestry. CHB, Han Chinese in Bejing, China. YRI, Yoruba in Ibadan, Nigeria. ASW, Americans of African Ancestry in South West, USA. n/a, not applicable.

Zang *et al.* (2001) first sequenced the *NR1I2* gene in 150 Caucasian and 22 African American patients and reported 38 SNPs, including five 5'-upstream SNPs, two 5'-untranslated region (UTR) SNPs, three synonymous and three non-synonymous SNPs. One non-synonymous SNP, rs12721613 (P27S), was non-polymorphic in the Caucasian cohort and another SNP, rs12721607 (G36R), was observed at a higher frequency in African Americans than in Caucasians. In the 1000 Genomes data, P27S was detected in the Yoruban (YRI) and African American South West (ASW) populations (MAF = 0.188 and 0.107), but was non-polymorphic in the European Caucasian (CEU) and Chinese Han (CHB) population. G36R was present in the ASW population only (MAF = 0.016). The SNP rs12721608 (R122Q), polymorphic in only Caucasian individuals, was associated with a decrease in CYP3A4-promoter activation by PXR following *in vitro* PXR activity assays using reporter gene constructs.

King *et al.* (2007) sequenced the *NR1I2* gene including the 5'-UTR and 3'-UTR in 120 individuals of Caucasian, African American and Asian ancestry and identified 52 SNPs, of which 14 (27%) were only polymorphic in the African American cohort. These SNPs included four non-synonymous SNPs, three of which were unique to the African American cohort. One SNP in the

5'-UTR, rs1523127 (6709C>A), was associated with altered PXR activity. In this study, patients with the rs1523127 A-allele presented a significant increase in the PXR-induction of both *CYP3A4* and *ABCB1* transcription (P = 0.04 for both genes) (King *et al*. 2007).

The sequencing of the 16 kb promoter region and intron 1 of *NR1I2* by Lamba *et al.* (2008) in 24 individuals resulted in the identification of 89 SNPs. Analysis of CYP3A4-activity in the presence of rifampicin identified five important SNPs. The T-alleles of two of these SNPs, rs13085558 (18885T>C) and rs2472679 (30120A>T), were significantly associated with decreased CYP3A4 induction by rifampicin (P = 0.010 and 0.023 respectively). Two other SNPs, rs2472677 (24087C>T) and rs1523130 (5177T>C), were significantly associated with increased basal CYP3A4 activity (P = 0.006 and 0.037) and one other SNP, rs3814055 (5705C>T), was significantly associated with increased induction of CYP3A4 by rifampicin (P = 0.046). After analysis of hepatic PXR expression, rs3814055, was also reported to be associated with decreased PXR expression in male individuals (P = 0.029).

Even though these five studies present some overlap in SNP identification and selection, none of the significant functional findings was replicated in any of the studies. For example, the non-synonymous SNP rs72551372 (V140M) was detected by both Hustert *et al.* (2001) and Svärd *et al* (2010), however, only the latter study was able to find significant statistical evidence of the influence of the SNP on PXR activity.

2.2.4.2. The role of NR1I2 variants in EFV metabolism and treatment outcome

There is little evidence that directly links *NR1I2* genetic variation with EFV pharmacokinetics and treatment outcome. The T-allele of the rs2472677 SNP, as discussed above, was associated with higher CYP3A4 activity. In the study by Wyen *et al.* (2011), this SNP bordered on significance in EFV treatment outcome, as patients with the T-allele experienced a higher instance of EFV discontinuation (P = 0.072). It can therefore be hypothesised that an increase in CYP3A4 activity may result in a more rapid breakdown of EFV in the body and subsequent sub-therapeutic EFV levels.

Patients with the G/G-genotype for the rs3732356 (34783G>T) SNP experienced the lowest instance of EFV discontinuation after one year (0%), compared to patients with the T/G genotype (24%) and the T/T genotype (21%) (Swart *et al*., 2012). However, no indication of statistical significance was given, and the sample size of patients with the G/G-genotype was small (n = 9).

Table 4 lists the MAF of the rs2472677 and rs3732356 SNPs. SNP rs2472677 is well distributed across all the ethnic groups, however, the minor allele for the African and African American populations differ from the Caucasian and Asian populations. SNP rs3732356 occurs at a low MAF in Caucasian and Asian populations.

Table 4: Population and functional details of *NR1I2* SNPs associated with EFV SNPs in populations of Caucasian (Cau), Asian (Asn), African (Afr) and African American (AA) ethnicities.

Location	rs number	Genomic identifier	MAF				Relevance	Reference
			Cau (CEU)	Asn (CHB)	Afr (YRI)	AA (ASW)		
Intron 1	rs2472677	24087C>T	0.359 (C)	0.371 (C)	0.352 (T)	0.361 (T)	EFV discontinuation	Wyen *et al.*, 2011; Swart *et al.*, 2012
Intron 3	rs3732356	34783G>T	0.071	0.072	0.295	0.279	EFV discontinuation	Swart *et al.*, 2012

CEU, CHB, YRI and ASW, represent four respective ethnic groups that form part of the 1000 Genomes Project. CEU, Residents in Utah, USA, of Northern and Western European ancestry. CHB, Han Chinese in Bejing, China. YRI, Yoruba in Ibadan, Nigeria. ASW, Americans of African Ancestry in South West, USA.

2.3. The Cape Mixed Ancestry population

The ethnic group known as Cape Mixed Ancestry (CMA), or South African Coloured, forms 9% of the South African population, and 54% of the Western Cape population (De Wit *et al.*, 2010; Statistics South Africa, 2013). The use of individuals from an admixed population, such as CMA, in pharmacological association studies, holds certain advantages and disadvantages. Depending on the length of time since the original ancestries were combined, admixture results in the inheritance of longer sequence regions in linkage disequilibrium (LD), which reduces the number of variants or markers needed to cover a certain region to analyse effects on phenotype (Baye & Wilke, 2010). However, an admixed population presents a higher risk of stratification and the presence of genetically similar sub-groups within the study cohort, which can lead to results that are mistakenly reported as significant (Baye & Wilke, 2010). Knowledge of the ancestries that contribute to the genetic composition of a specific admixed cohort is therefore advantageous in association studies.

The genetic profile of an admixed population can be analysed using microsatellites, SNPs and indels (insertions and deletions) to form structural and variance patterns that differ among

ancestries, as graphically illustrated in the CMA population in figure 3. The prediction power of these patterns is enhanced through analysis of LD among variants (De Wit *et al.*, 2010).

Two studies examined the major ancestries contributing to CMA genetic composition. Tishkoff *et al.* (2009) reported genetic contributions to CMA by Khoisan ancestry (25%), Niger-Kordofanian (N-K) ancestry (19%), which is shared by groups from Central and Western Africa, Indian ancestry (20%) and European ancestry (19%) (Figure 3A). De Wit *et al.* (2010), accounting for LD, reported genetic components from Khoisan (32%), African Bantu (36%), European Caucasian (21%) and Indian (11%) ancestries (Figure 3B) in the CMA population. The De Wit *et al.* (2010) study reports that two geographically separate CMA cohorts are genetically compatible; however, further studies are needed to investigate the possible genetic differences in CMA sub-groups.

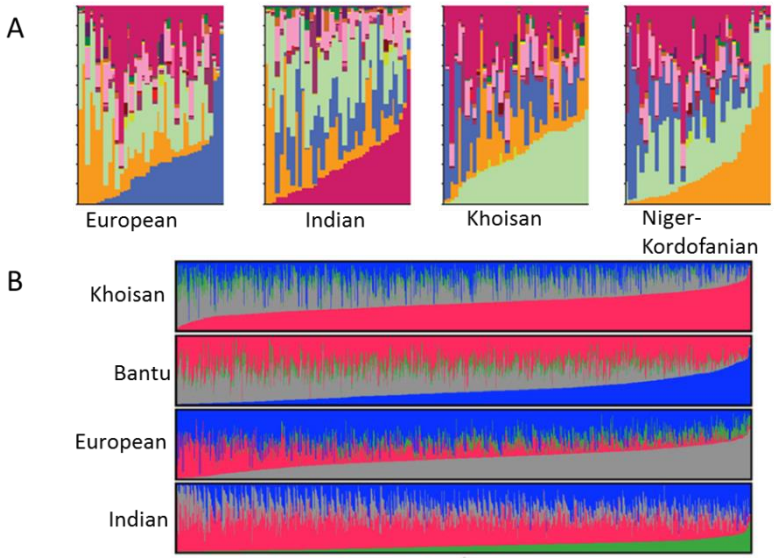


Figure 3: Structural and genetic patterns that contribute to the genetic contents of the CMA population in two studies, arranged in four different ways according to the major contributing ancestries.
Each individual is represented as one bar and divided into four colours according to the proportion of European, Indian, Khoisan and Niger-Kordofanian ancestry present. A. Adapted from Tishkoff *et al.* 2009, with publisher's permission. B. Adapted with De Wit *et al.*, 2010, with publisher's permission.

2.4. The Xhosa population

The Xhosa-speaking population constitutes 17.6% of the South African population and 24.7% of the Western Cape population (Statistics South Africa, 2012). Despite the large population size, limited population genetic information was available for the Xhosa population until 2009, when a

comprehensive study investigating the genetic variation among African populations, highlighted the need for further analysis of these populations (Tishkoff *et al*., 2009).

When analysing the genetic composition of the Xhosa population, which form part of the Southern African Bantu population, the main contributing populations are from N-K (populations from Central and Western Africa, including the Yoruban population), and Khoisan populations (Tishkoff *et al*., 2009). The genetic constituency of the Xhosa population thus forms a separate group from the Khoisan and N-K populations, as presented through principal component analysis using 174,272 SNPs (Figure 4) (Schuster *et al*., 2010).

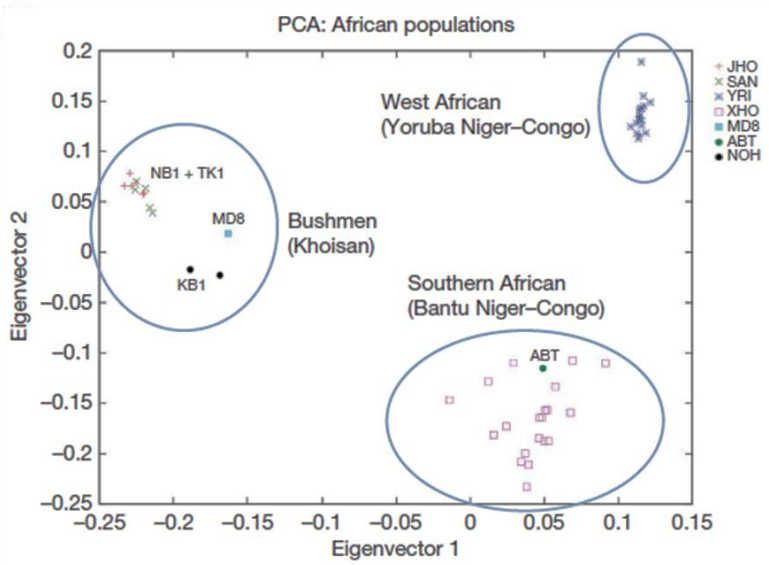


Figure 4: Representation of the genetic structure of the Xhosa population to illustrate their separation from West African and Khoisan populations.
Data obtained from principal component analysis using approximately 170,000 SNPs. The Xhosa population forms part of the Southern African (Bantu Niger Congo) population group. JHO, MD8 and NOH, Khoisan individuals. SAN, Bushmen. YRI, Yoruban individuals. XHO, Xhosa individuals. ABT, Xhosa/Tshwana Bantu individual. Taken from: Schuster *et al*., 2010, with publisher's permission.

2.5. Significance of the study

The validity of this study is proven through a summary of some of the discussion points in the literature review above. These points illustrate how this study can contribute to the understanding of the scope of genetic variance influencing EFV pharmacokinetics and treatment outcome, as well as describing the central role that NRs CAR and PXR play in the complex system of xenobiotic metabolism.

- The vast number of people who are HIV-positive and enrolled for ART globally and in South Africa, reveal the gravity of this pandemic, both on a global and national scale. (Sections 2.1.1 and 2.1.4)
- The annual rapid increase of patients enrolled for ART indicate the necessity for effective management of ART on a global and national scale. (Section 2.1.1)
- Ineffective management of ART leads to viral resistance and treatment failure highlights the need to measure individual treatment outcome accurately and to establish a global consensus on the definition of good and poor treatment outcome. (Sections 2.1.2 and 2.1.3)
- The roles of *CYP1*-, *CYP2*-, and *CYP3*-genes in EFV metabolism have been the focus of several studies; however, a more thorough understanding of the activation and transcriptional regulation of the CYP-genes is needed. (Section 2.1.5.1)
- The overwhelming evidence for the influence of the *CYP2B6* 516G>T SNP, and its composite genotypes, on EFV metabolism warrants further investigation into other variants that cooperate with the 516T-allele. (Section 2.1.5.2)
- The ability of CAR and PXR to regulate the transcription of various genes involved in the breakdown of xenobiotic substances show that these NRs play an integral role in xenobiotic metabolism, including EFV metabolism. (Section 2.2.2)
- The promiscuous nature of the ligand binding domains of CAR and PXR reveals that the transcription of these NRs are influenced by various xenobiotic substances, which could ultimately influence the rate of EFV metabolism. (Section 2.2.2)
- The influence of intronic, exonic and upstream variants in *NR1I3* and *NR1I2* on gene transcription and protein activity demonstrated in previous studies warrant further analyses of the effects of variants in these genes on EFV metabolism and treatment outcome, especially in the diverse South African populations (Sections 2.2.3 and 2.2.4).

2.6. Research aim and objectives

This study aims to analyse variation in *NR1I3* and *NR1I2*, encoding CAR and PXR respectively, in female South African HIV-positive Xhosa and Cape Mixed Ancestry patients and, in conjunction with variations in *CYP2B6*, *CYP3A*, *CYP1A2* and *CYP2A6* to investigate correlations with EFV levels, treatment outcome, and adverse effects.

This aim will be achieved by the following objectives:

1. Identify target variants in *NR1I3* and *NR1I2* through bi-directional sequencing, data-mining of the HapMap Project and 1000 Genomes Project databases and a comprehensive literature search.
2. Genotype these target variants in 132 Xhosa and Mixed Ancestry HIV-infected female patients.
3. Utilise statistical analyses to search for correlations among the target variants, EFV levels in hair samples, and treatment outcome, by means of VL and the occurrence of adverse effects as data parameters.

CHAPTER 3

3. Materials and Methodology

The current study can be divided into four components, following each other chronologically: Sample and clinical data collection; target variant identification; target variant genotyping; and variant genotype analysis and statistical analysis (Figure 5). The project started with the recruitment of eligible study participants at the TC Newman HIV clinic in Paarl and the subsequent collection of hair and saliva specimens and clinical data. The hair specimens were used to determine long term EFV levels, and the saliva specimens were used to extract genomic DNA.

Following recruitment and sample collection, target variants were identified in *NR1I3* through bi-directional sequencing and in *NR1I2* through a survey of the HapMap and 1000 Genomes databases and literature. The identified target variants were then genotyped in the cohort of 79 Xhosa and 53 CMA HIV positive patients using an assortment of genotyping techniques. The genotyping results were then analysed in conjunction with measured EFV levels and clinical data in order to find any statistically significant correlations.

3.1. Study participants

A total of 136 patients were initially recruited for this study, but only 132, consisting of 79 patients of Xhosa (Xh) ethnicity and 53 patients of CMA ethnicity, were included for genetic studies. Ethnicity was verbally reported by the patients themselves. All patients were above the age of 18, female and HIV-positive. The patients were recruited at the TC Newman HIV Clinic in Paarl, Western Cape. All patients were prescribed an EFV-containing first-line ART regimen as directed by the South African Treatment Guidelines (National Department of Health, 2004) for a minimum period of three months. Other ARV drugs taken with efavirenz include d4T, lamivudine (3TC), tenofovir (TDF) and AZT. Consent were given by the patients for the participation in this study, and ethical approval was obtained by the Committee for Human Research, Stellenbosch University (Reference Number: N08/08/225) and the Department of Health (DOH), Western Cape.

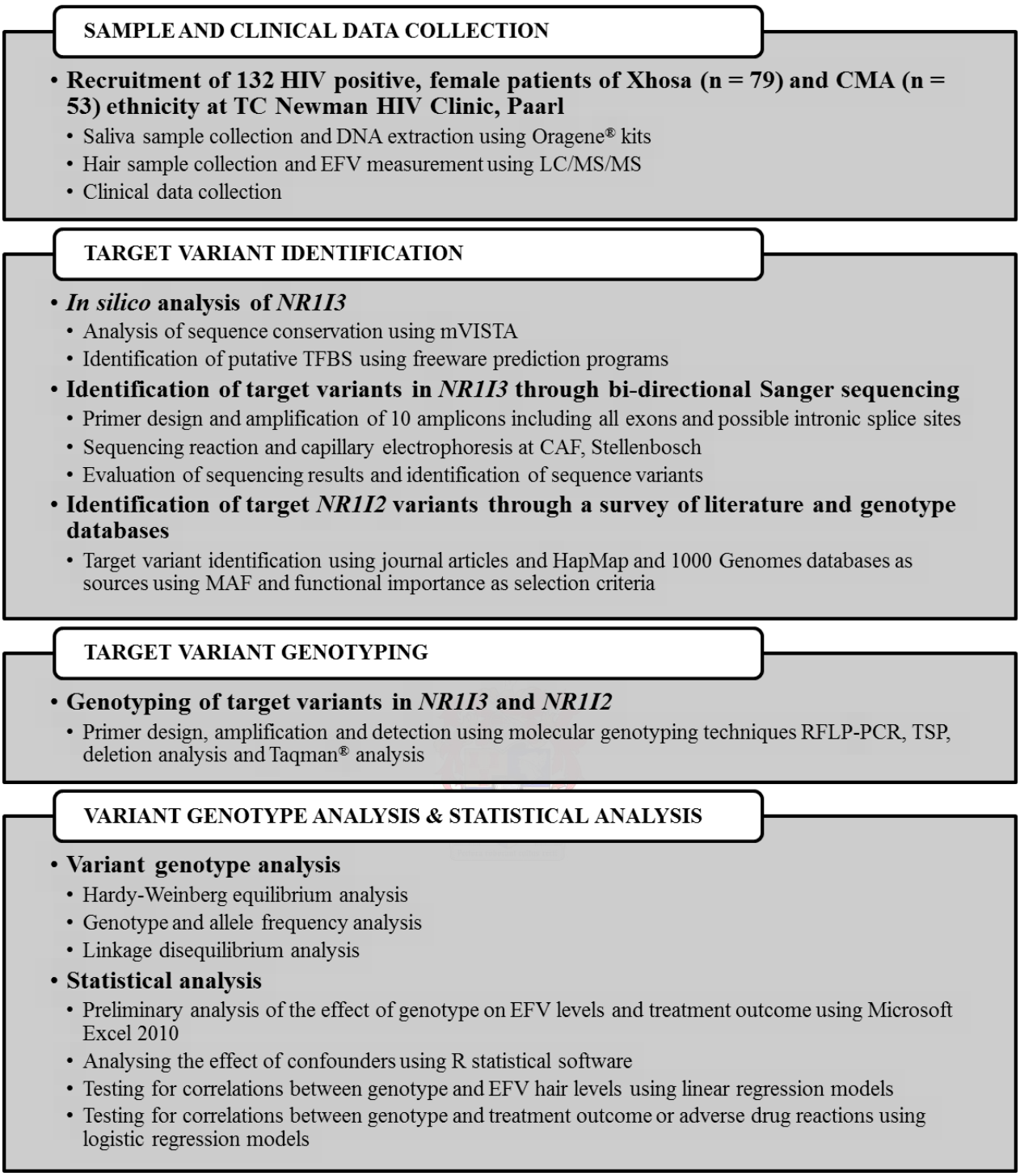


Figure 5: Flow diagram representing the chronological division of the current project into four sections, specifically, sample and clinical data collection, target variant identification, target variant genotyping and statistical analysis.

3.2. DNA sample collection and extraction

A minimum of 2 ml saliva were collected in Oragene® collection units (DNA Genotek Inc., Kanata, Canada) from each patient (Röhrich, 2012) and labelled with individual patient identification (ID) numbers. The samples were prepared for storage following the protocol provided by the supplier (Appendix A) and stored at -20°C. DNA was extracted as needed from the samples following the Oragene® protocol (Appendix A). The extracted DNA stock concentrations were measured for each sample using the NanoDrop® ND-100 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, USA). From the stock DNA, diluted aliquots with a concentration of 15 ng/µl were prepared with the addition of Sabax water (Adcock Ingram Pty. Ltd., Midrand, ZA).

3.3. Hair sample collection and EFV measurement

Hair samples were collected from each patient as reported in Röhrich (2012). Roughly 20 strands of hair were cut off close to the scalp at the back of the head from each patient and labelled with patient ID number and hair colour. Samples were covered in foil and sent to the Department of Clinical Pharmacy, University of California, San Francisco (UCSF) in the USA for analysis. The technique used for EFV measurement was liquid chromatography tandem mass spectrometry (LC/MS/MS), the protocol of which is described in Huang *et al.* (2008). The average weight of the hair samples were 2 mg and EFV levels were measured in ng EFV per mg of hair sample (ng/mg).

3.4. Clinical data collection

The TC Newman HIV clinic supplied clinical data for 131 patients from clinical folders and from an electronic database. Basic personal information was obtained, consisting of date of birth, hospital ID, alcohol consumption and smoking habits. Important dates were recorded including date of HIV diagnosis, start of ART, start of EFV-based ART and date of sample collection. Clinical baseline information was obtained at the start of treatment, which included height, weight, WHO disease stage and CD4 count. The level of adherence was described through a range of pill-count percentages: When 70 – 100% of medication was taken it was considered good adherence, acceptable adherence was considered as 50 – 70% and bad adherence was considered as 0 – 50%. The list of adverse drug reactions obtained consisted of CNS-ADRs (headaches, dizziness, dreams, insomnia, depression and psychosis), liver toxicity, lactic acidosis, anaemia, lipodystrophy, peripheral neuropathy and work impairment. Adverse drug reactions

were reported through patient verbal communication, clinical examination and blood tests. Records of patient default of treatment, changes in regimen and the accompanying reasons were obtained. Measurements of CD4 count, VL and weight were recorded every six months. The list of additional medication collected consisted of medication for TB (primarily rifampicin), diabetes, hypertension, asthma, hypercholesterolemia, epilepsy, as well as anticoagulants and chemotherapeutic agents. Additionally, the genotypes of patients for the *CYP2B6*6* allele was available from a previous study (Röhrich, 2012).

3.5. *In silico* analysis of *NR1I3*

In order to identify upstream amplicons (UAs) in the promoter and 5'-upstream area of the *NR1I3* gene, two strategies were followed. One strategy followed the claim that regions of high sequence conservation among orthologs of the human *NR1I3* gene could be important to gene function. Therefore the promoter region of *NR1I3* was searched for conserved regions. Another strategy followed the idea that possible transcription factor binding sites (TFBSs) in the gene promoter could be important for transcriptional regulation of the gene. Therefore, the promoter area of *NR1I3* was scanned for putative TFBS.

3.5.1. Analysis of sequence conservation using mVISTA

Areas of genetic conservation in the *NR1I3* promoter and 5'-upstream region were analysed using mVISTA (VISTA Tools for Comparative Genomics, Frazer, *et al.*, 2004). Orthologs for the human *NR1I3* gene was identified using the Ensembl Genomes Browser and HomoloGene. The orthologs selected and their respective percentage identities to the human *NR1I3* gene are listed in Table 5. Sequences for the 5kb-upstream region, the first two exons and the first intron for each gene were obtained from the Ensembl Genomes Browser. The mVISTA software was used to find areas of sequence conservation among these homologous genes.

Table 5: List of orthologs used in the sequence conservation analysis using mVISTA and their percentage identity to the *NR1I3* human sequence.

Gene	Species	Ensembl identifier	% identity
NR1I3	*Homo sapiens* (human)	ENST00000367983.4	100
NR1I3_PANTR	*Pan troglodytes* (chimpanzee)	ENSPTRT00000002884	100
NR1I3	*Gorilla gorilla* (gorilla)	ENSGGOT00000005750	99
A2T809_PONPY	*Pongo abelii* (orangutan)	ENSPPYT00000000729	98
Q38IV5_CANFA	*Canis familiaris* (dog)	ENSCAFT00000014526	85
Q2KIF4_BOVIN	*Bos Taurus* (cow)	ENSBTAT00000012145	84
NR1I3	*Rattus norvegicus* (rat)	ENSRNOT00000049873	77
NR1I3	*Mus musculus* (mouse)	ENSMUST00000005820	73
NR1I3	*Felis catus* (cat)	ENSFCAT00000000496	71

3.5.2. Identification of putative transcription factor binding sites using freeware prediction programs

Further analysis of the promoter area of *NR1I3* prompted the search for putative TFBSs. Free prediction software programs were used to find putative TFBSs in the 5 kb upstream area, exon 1 and intron 1 of *NR1I3*. *NR1I3* gene and transcript sequences were taken from the Ensembl Genomes Browser (Ensembl identifiers ENSG00000143257, ENST00000367983.4) and analysed using the default settings in six programs, Match, AliBaba, Patch, P-Match, Jaspar and rVISTA (Grabe, 2002; Loots *et al*., 2002; Kel, 2003; Sandelin *et al*., 2004; Chekmenev *et al*., 2005). Results were compiled in Microsoft Office Excel 2010 and compared to functional TFBSs in *NR1I3* identified in literature.

3.6. Identification of target variants in *NR1I3* through bi-directional Sanger sequencing

3.6.1. Primer design

Sequences for a total of 10 amplicons were obtained from the Ensembl Genomes Browser (Ensembl identifier ENST00000367983.4). The amplicons include the areas encoding the DNA binding domain (DBD) and parts of the ligand binding domain (LBD) of *NR1I3*. Figure 6 presents the location of the primer sets, as well as the regions amplified in *NR1I3*. Primers were either designed using PrimerQuest (Integrated DNA Technologies Inc., Coralville, USA) or obtained from literature (Ikeda *et al.*, 2003; Thompson *et al.*, 2005). Evaluation of primer melting temperature (T_M), GC-content, dimer-formation and hairpin-formation was performed using OligoAnalyzer (Integrated DNA Technologies Inc., Coralville, USA) and evaluation of primer

specificity was performed using primer-BLAST (pBLAST) (Ye *et al.*, 2012). Primers for amplifying the exonic regions were designed to allow a flanking region of +/- 50 bp both sides of the exons where possible in order to incorporate potential splice sites. All primers used for the amplification of the target regions for bi-directional sequencing are listed in Appendix B. Primers were supplied by Integrated DNA Technologies (IDT) (Integrated DNA Technologies Inc., Coralville, USA).

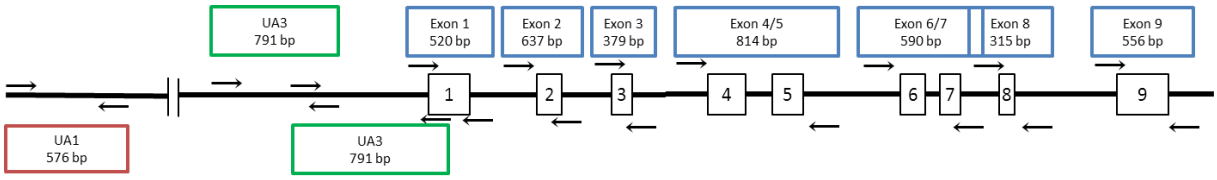


Figure 6: Regions of *NR1I3* amplified for sequencing, consisting of 10 amplicons.
UA1 (red box) is located 4.2 kb from the start of exon 1. UA2 and UA3 (green boxes) are located just upstream of exon 1. The region in the blue boxes includes the 5'-UTR present in exon 1 and a part of exon 2, and also the rest of the coding region (a total region of 8.9 kb).

3.6.2. PCR amplification

Following optimisation, all 10 amplicons in *NR1I3* were amplified using 15 ng of each diluted DNA sample for 15 Xhosa individuals and 15 CMA individuals, resulting in a total of 300 polymerase chain reactions (PCRs). PCR amplification was performed on either GeneAmp 9700 or GeneAmp2700 thermocyclers (Life Technologies™ Corporation, Carlsbad, USA) using the following general cycle conditions: 94°C for three min, one cycle; 94°C for 15 sec, T_A for 15 sec, 72°C for 30 sec, 35 cycles; 72°C for five min, one cycle. Annealing temperatures (T_A) of 64°C (for UA1 and UA3), 60°C (for Exon 1 and Exon 2), 54°C (for Exon 3), 58°C (for Exon 4/5) and 62°C (for Exon 6/7) were used.

Exons 8 and 9 were amplified through hotstart-PCR with the following cycle conditions: 94°C for three min, one cycle; 94°C for 15 sec, T_1 for 15 sec, 72°C for 30 sec, 10 cycles; 94°C for 15 sec, T_2 for 15 sec, 72°C for 15 sec, 12 cycles; 72°C for five min. Annealing temperatures (T_1 and T_2) of 58°C and 54°C (Exon 8); and 66°C and 60°C (Exon 9) were used.

The total reaction mixture of 25 µl used for all amplicons consisted of 1x ammonium (NH_4) reaction buffer, 1.5 mM magnesium chloride ($MgCl_2$), 0.4 mM deoxynucleotide triphosphate molecules (dNTPs), the primers discussed above (forward and reverse primers for each PCR), 0.5 U BIOTAQ™ DNA polymerase, and Sabax water. All reagents in the reaction mixture were

supplied by Bioline™ (London, UK), except for the primers (Integrated DNA Technologies Inc., Coralville, USA) and the Sabax water (Adcock Ingram Pty. Ltd., Midrand, ZA). The amount of each primer used was 10 pmol from a 10 pmol/µl stock solution for all amplicons except UA1 and Exon 4/5, where 8 pmol of each primer was used.

Electrophoresis was performed for the amplified PCR products on 1.0% (w/v) agarose gels (Lonza Group Ltd., Rockland, USA) prepared with a 1x TBE buffer solution (Appendix C) and stained with 2.5 µl ethidium bromide (Sigma-Aldrich Co. LLC, St. Louis, USA). A volume of 5 µl of PCR product, mixed with 5 µl cresol loading dye (Appendix C) was loaded onto the agarose gel, along with a Hyperladder IV (Bioline™, London, UK) molecular marker. The gels were exposed to an electric current for 30-45 min at 120 V and visualised under ultraviolet (UV) light at a wavelength of A_{260} nm using the MultiGenius bio-imaging system (Syngene, Synoptics Ltd., Cambridge, UK).

3.6.3. Sequencing reaction and capillary electrophoresis

PCR products were subjected to a cleaning protocol (SureClean, Bioline™, London, UK) (Appendix A) and cleaned product concentrations were determined using the NanoDrop® ND-100 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, USA). Bi-directional sequencing reactions were performed using 3-20 ng of the cleaned PCR product, 3.3 pmol of a 1.1 pmol/µl primer dilution (forward and reverse primers were used in separate reactions), 2.7 µl of Half-Dye Mix (Bioline™, London, UK), 1.3 µl of BigDye® Terminator v1.1 Ready Reaction Mix (Life Technologies™ Corporation, Carlsbad, USA) in a final reaction volume of 10 µl. The sequencing reaction was performed on a GeneAmp 9700 thermocycler (Life Technologies™ Corporation, Carlsbad, USA) and the following cycle conditions were used: 94°C for five min, 1 cycle; 94°C for 10 sec, 55°C for 10 sec, 60°C for four min, 25 cycles. Following the sequencing reaction, 1 µl of a 2.2% (w/v) sodium dodecyl sulphate (SDS) solution (Appendix C) was added to the products, which were then subjected to a thermal cycle of 98°C for five min and 25°C for 10 min. Following SDS clean-up, the products were sent to the DNA Sequencing Central Analytical Facility (CAF, Stellenbosch, ZA) for visualisation through capillary electrophoresis using a 3130XI Genetic Analyzer (Life Technologies™ Corporation, Carlsbad, USA).

3.6.4. Evaluation of sequencing results and identification of sequence variants

Chromatograms for each patient were visualised in BioEdit Sequence Alignment Editor (Hall, 1999) and aligned to reference sequences from the Ensembl Genomes Browser. Only variants

that were present in the sequencing results of both directions were accepted and used in further analysis. The identified variations were then evaluated in order to select target variants for genotyping. Priority was given to variants with functional importance in literature, the disruption of putative TFBSs as predicted by AliBaba (Grabe, 2002), the possible effect of non-synonymous variants on protein activity as predicted by SIFT and PolyPhen (Ramensky *et al*., 2002; Ng, 2003), and the possible effect of variants on the folding of single-strand (ss) DNA and mRNA as predicted by mFOLD (Zuker, 2003).

3.7. Identification of target *NR1I2* variants through a survey of literature and genotype databases

During this study, online accessible databases for the HapMap and 1000 Genomes projects were established (The International HapMap 3 Consortium, 2010; The 1000 Genomes Consortium, 2012). These databases are easy to access, have a large amount of variation data available and use diverse populations (including African populations). Therefore it was decided to use these databases, in combination with a literature analysis, to find target variants for genotyping in the *NR1I2* gene. Through the use of databases PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and ScienceDirect (http://sciencedirect.com) and the search engine Google Scholar (http://scholar.google.co.za), suitable research journal articles were identified which focussed on *NR1I2* SNP population frequencies, and their impact on drug pharmacokinetics as well as other medically relevant phenotypes. Functional importance and allele frequency were used as benchmarks for selecting variants. A minimum MAF of 0.05 was considered as a cut-off value for selecting variants, with priority given to variants with a high MAF in African or African American populations. Functional characteristics of variants that were considered included the 1) disruption of putative or confirmed TFBS, 2) the influence on the transcription and activity of PXR and downstream proteins, and 3) influence on phenotypic outcomes including drug levels, clinical disease outcome and clinical treatment outcome.

3.8. Genotyping of target variants in *NR1I3* and *NR1I2*

A total of 19 variants were genotyped in the study cohort: Nine variants from *NR1I3*, nine variants from *NR1I2* and one short deletion from *NR1I2*. Various genotyping techniques were used, namely PCR restriction fragment length polymorphism assays (PCR-RFLP) (14 SNPs), temperature switch PCR assays (TSP) (Tabone *et al*., 2009) (three SNPs) and Taqman® SNP Genotyping assays (Life Technologies™ Corporation, Carlsbad, USA) (one SNP). A 6-bp-deletion (rs3842689) was characterised through PCR amplification and subsequent visualisation

of fragment size differences. The specific characteristics (T_A, $MgCl_2$ concentration and primer concentration) for each SNP assay are listed in Table 6.

Table 6: PCR cycle characteristics for genotyping assays for all SNPs, consisting of annealing temperature, $MgCl_2$ concentration and the amount of primer used per reaction.

Gene	SNP	Genotyping assay	T_A or T_1, T_2, T_3 (°C)	[$MgCl_2$] in mM	Primer amount in pmol[1] or µM[2]
NR1I3	rs55802895	PCR-RFLP	60	1.5	10
NR1I3	rs35709078	PCR-RFLP	54	1.5	10
NR1I3	rs3003596	PCR-RFLP	60	1.5	10
NR1I3	rs2502815	TSP	60, 47, 53	1.5	0.1, 0.5
NR1I3	rs35205211	PCR-RFLP	58	1.5	8
NR1I3	rs34161743	PCR-RFLP	58	1.5	8
NR1I3	rs2307424	TSP	64, 51, 53	2.0	0.1, 0.4
NR1I3	rs2307420	PCR-RFLP	60	1.5	10
NR1I3	rs113800307	PCR-RFLP	62	1.5	10
NR1I2	rs1523130	Taqman® analysis	N/A	N/A	N/A
NR1I2	rs3814055	PCR-RFLP	60[3], 55[4]	1.5	10
NR1I2	rs1523128	PCR-RFLP	60	1.5	10
NR1I2	rs1523127	TSP	60, 43, 53	2.0	0.1, 0.5
NR1I2	rs3842689	6-bp deletion	60	1.5	8
NR1I2	rs2472677	PCR-RFLP	60	1.0	10
NR1I2	rs12721613 (*PXR*2*)	PCR-RFLP	60	1.5	10
NR1I2	rs1464603	PCR-RFLP	60	1.5	10
NR1I2	rs1464602	PCR-RFLP	60	1.5	10
NR1I2	rs2276707	PCR-RFLP	60	1.5	10

[1]For PCR-RFLP and deletion assays: Primer amount as taken from a 10 pmol/µ stock solution. [2]For TSP assays: LS primer TM and NLS primer TM. [3]TM for the first amplification of the assay. [4]TM for the nested amplification of the assay.

3.8.1. Primer design

Primers used for PCR-RFLP, TSP or deletion assays for *NR1I3* and nine variants are given in Appendix B. Primers were either designed with the use of PrimerQuest (Integrated DNA Technologies Inc., Coralville, USA), or taken from literature (Ikeda *et al.*, 2003; Thompson *et al.*, 2005; Andrews *et al.*, 2010), and analysed using OligoAnalyzer (Integrated DNA Technologies Inc., Coralville, USA). The template sequences used for primer design, ENST00000367983.4 (*NR1I3*) and ENST00000393716 (*NR1I2*), were obtained from the Ensembl Genomes Browser. All primers were supplied by IDT.

Primers that best fit the following criteria were selected for the PCR-RFLP assays and deletion-assay: Primer size between 22 bp and 26 bp, a T_M between 55°C and 65°C, GC-content of 50% to 60%, minimum primer dimer formation and minimum hairpin formation. Primer specificity was analysed through pBLAST (Ye *et al.*, 2012). The assay for one SNP, rs3814055, required a nested PCR assay and primers were designed accordingly.

For the TSP-assays, two sets of primers were designed for each SNP, a locus-specific (LS) primer set and a nested locus-specific (NLS) primer set with T_M the key difference between the two sets. The LS primers required a T_M between 60°C and 65°C and the resulting amplicon size exceeded 400 bp. The NLS primers required a T_M between 43°C and 47°C and the resulting amplicon size was between 200 bp and 300 bp. A non-complementary 5'-tail was added to all NLS-primers in order to increase the T_M once annealing has taken place.

3.8.2. Restriction fragment length polymorphism analysis

PCRs for PCR-RFLP assays were performed on either GeneAmp 9700 or GeneAmp 2700 thermocyclers using a reactant mix with a total volume of 25 µl consisting of 1x NH4 reaction buffer, $MgCl_2$, 0.4 mM dNTPs, forward and reverse primers, 0.5 U BIOTAQ™ DNA polymerase and Sabax water. Each reaction contained 15 ng of DNA (1 µl). The general cycle conditions used for amplification were: 94°C for three min, one cycle; 94°C for 15 sec, T_A for 15 sec, 72°C for 30 sec, 35 cycles; 72°C for five min, one cycle.

Ampified DNA products were visualised through electrophoresis using ethidium bromide-stained 1.0% agarose gels prepared with 1X TBE buffer and a voltage of 120 V for 30 minutes (see Appendix C). A total loading volume of 10 µl consisting of 5 µl PCR product and 5 µl cresol loading dye for each sample was loaded onto the agarose gel and Hyperladder V (Bioline™, London, UK) was used as the molecular marker. Following electrophoresis, the gels were visualised with the MultiGenius bioimaging system.

PCR products were subjected to restriction enzyme digestion in a waterbath at specific temperatures as specified in Table 7, using enzymes supplied by New England Biolabs®, Inc. (NEB), Ipswitch, USA and Promega Corporation, Fitchburg, USA. The reactant mix for digestion consisted of 5µl PCR product, 1x reaction buffer, enzyme and Sabax water and amounted to a total volume of 20 µl. Table 7 lists the enzymes and amount of enzyme used, buffers, incubation time and additional reagents used for each SNP.

Table 7: Enzymes used for all SNPs genotyped using PCR-RFLP analysis.

SNP	Enzyme	Buffer	Supplier	Amount in units (u)	Incubation temperature (°C) and time	Additional reactants
rs113800307	*Nco I*	NEBuffer 3	NEB®	3 u	37°C, 1 hr	n/a
rs2307420	*Pst I*	NEBuffer 3	NEB®	5 u	37°C, overnight	100 µg/ml BSA
rs34161743	*MspA I*	NEBuffer 4	NEB®	3 u	37°C, 1-4 hrs	100 µg/ml BSA
rs35205211	*Hha I*	Buffer C	Promega	3 u	37°C, 1-4 hrs	n/a
rs3003596	*Hpy188 I*	NEBuffer 4	NEB®	3 u	37°C, overnight	100 µg/ml BSA
rs35709078	*Bfa I*	NEBuffer 4	NEB®	3 u	37°C, 1-4 hrs	n/a
rs55802895	*Hph I*	NEBuffer 4	NEB®	3 u	37°C, 1 hr	n/a
rs3814055	*Mbo I*	NEBuffer 2	NEB®	3 u	37°C, 1 hr	n/a
rs1523128	*Hha I*	NEBuffer 4	NEB®	3 u	37°C, 1 hr	100 µg/ml BSA
rs2472677	*Hpy188 I*	NEBuffer 4	NEB®	3 u	37°C, 1 hr	n/a
rs12721613 (*PXR*2*)	*Bmr I*	NEBuffer 2.1	NEB®	3 u	37°C, 1 h	n/a
rs1464603	*BsmB I*	NEBuffer 3	NEB®	5 u	37°C, overnight	n/a
rs1464602	*BsmF I*	NEBuffer 4	NEB®	3 u	65°C, 1 hr	100 µg/ml BSA
rs2276707	*Bcc I*	NEBuffer 1	NEB®	3 u	37°C, 1 hr	n/a

NEB®, New England Biolabs®. BSA, bovine serum albumin. hr(s) hour(s).

For SNP rs2276707, digested products were visualised using 15% (w/v) polyacrylamide gels (see Appendix C). All polyacrylamide gels were prepared using 5X TBE Buffer, prepared from a 10X TBE stock solution (see Appendix C). A total loading volume of 15 µl consisted of 10 µl digested PCR product and 5 µl Bluejuice loading dye (Appendix C) for each sample. Additionally, a molecular marker (Hyperladder V, Bioline™, London, UK) was loaded onto the gel. Electrophoresis was performed using a Mighty Small II S250 unit (Amersham Biosciences, GE Healthcare, Buckinghamshire, UK) at 150 V for 90 minutes. Following electrophoresis, the gels were stained using ethidium bromide and visualised with the MultiGenius bio-imaging system. For the remaining 13 SNPs, digested procucts were visualised through electrophoresis of agarose gels. For each digested product, 7 µl digested product and 7 µl cresol loading dye was loaded onto a 3% (w/v) ethidium-bromide stained agarose gel, along with Hyperladder IV and V (Bioline™, London, UK) as molecular markers. Gel electrophoresis was performed at a voltage of 80 V for +/- two hours, then visualised under UV light.

3.8.3. Temperature switch PCR analysis

All PCR-reactions for TSP genotyping assays were performed on a Veriti™ thermocycler (Life Technologies™ Corporation, Carlsbad, USA) with the following cycle conditions: 94°C for three min, 1 cycle; 94°C for 30 sec, T_1 for 30 sec, 72°C for 1 min, 15 cycles; 94°C for 10 sec, T_2 for 30

sec, five cycles; 94°C for 10 sec, T_3 for 30 sec, 72°C for five sec, 15 cycles; 72°C for five min, one cycle. The reaction mix for each TSP reaction consisted of 1x NH4 reaction buffer, $MgCl_2$, 0.4 mM dNTPs, forward and reverse primers, 100 ng/µl bovine serum albumin (BSA) and 0.15 U BIOTAQ™ DNA polymerase with a final total volume of 8 µl. All reagents were supplied by Bioline™ (London, UK), except BSA (New England Biolabs®, Inc. (NEB), Ipswitch, USA) and Sabax water (Adcock Ingram Pty. Ltd., Midrand, ZA). Table 6 summarises the unique PCR-characteristics of each TSP genotyping assay.

Fragment size differences were visualised with electrophoresis of 1.5% (w/v) agarose gels prepared with 1X TBE at a voltage of 100 V for 45 minutes, and stained with ethidium bromide. An amount of 5 µl cresol solution was added to each TSP-product, after which 10 µl of the cresol-product mix was loaded onto the agarose gel along with a Hyperladder IV (Bioline™, London, UK) molecular marker. The gels were visualised with the MultiGenius bioimaging system.

3.8.4. Deletion analysis

A 6 bp deletion, rs3842689, was genotyped through PCR amplification of the sequence flanking the deletion (Figure 7) and the subsequent visualisation of different fragment sizes of the resulting amplicon. The reactant mix with a total volume of 25 µl consisted of 15 ng of a 15 ng/µl DNA sample dilution, 1x NH4 reaction buffer, 1.5 mM $MgCl_2$, 0.4 mM dNTPs, 8 pmol of 10 pmol/µl forward and reverse primer dilutions, and 0.5 u BIOTAQ™. The fragment was amplified using a PCR cycle of: 94°C for three min, one cycle; 94°C for 15 sec, 60°C for 15 sec, 72°C for 30 sec, 30 cycles, 72°C for five min, one cycle.

Figure 7: Deletion rs3842689 (bold and in brackets) and its flanking sequence (in the 5' to 3' direction).
The arrows indicate the annealing sites of the forward and reverse primers.

Resulting amplicons were visualised on a 15% polyacrylamide gel prepared with 5X TBE buffer (Appendix C). A Mighty Small II S250 unit (Amersham Biosciences, GE Healthcare, Buckinghamshire, UK) was used for electrophoresis. A total mixture of 15 µl consisting of 10 µl

PCR product and 5 µl Bluejuice (Appendix C) was loaded onto the gel, followed by electrophoresis at 120 V for 90 minutes. Gels were stained with ethidium bromide and fragments were visualised under UV light using the MultiGenius bioimaging system.

3.8.5. Taqman® analysis

One SNP, rs1523130, was genotyped using TaqMan® Drug Metabolism Genotyping Assay (assay number *C___9152783_20*, Life Technologies™ Corporation, Carlsbad, USA). A reaction mix with a total volume of 10 µl consisting of 15 ng sample DNA, 1x Taqman® Genotyping Master Mix, 1x TaqMan® Drug Metabolism Genotyping Assay and Sabax water was pipetted onto a MicroAmp® Fast Optical 96-well plate (Life Technologies™ Corporation, Carlsbad, USA) and covered with a MicroAmp® Optical Adhesive Film. After centrifuging, PCR amplification and probe incorporation was performed with the StepOne™ Real-Time PCR System (Life Technologies™ Corporation, Carlsbad, USA). The PCR cycle program was set using StepOne™ Software v2.2.2 and consisted of 60°C for 30 sec, 1 cycle; 95°C for 10 min, 1 cycle; 95°C for 15 sec and 60°C for 1 min, 60 cycles. Results were analysed using StepOne™ Software.

3.9. Variant analysis

3.9.1. Genotype and allele frequency analysis

Allele frequencies for all variants were compared between the Xhosa cohort and the CMA cohort. The variant MAFs for both cohorts were also compared to MAFs of five population groups from the 1000 Genomes Project and one population group from the HapMap Project (The International HapMap 3 Consortium, 2010; The 1000 Genomes Consortium, 2012). Descriptions for all populations used are listed in Table 8. Genotype and allele information of all variants were collected from the 1000 Genomes Browser in Ensembl and from the HapMap database. Contingency tables (http://www.vassarstats.net) were used to search for significant differences in genotype and allele frequency between the study cohorts and the 1000 Genome and HapMap population groups. MAF information for the HapMap and 1000 Genomes populations were not available for variants rs113800307 and rs3842689, resulting in the omission of these two variants from the analysis.

Table 8: Description of HapMap (HM) and 1000 Genomes (1000 G) populations.

Population Name	Database	Description
LWK	1000 G	Luhya in Webuye, Kenya
YRI	1000 G	Yoruba in Ibadan, Nigeria
ASW	1000 G	Americans of African Ancestry in South West, USA
CEU	1000 G	Utah Residents (CEPH) with Northern and Western European ancestry
CHB	1000 G	Han Chinese in Bejing, China
GIH	HM	Gujarati Indian from Houston, Texas

Adapted from: http://www.1000genomes.org/faq/which-populations-are-part-your-study.

3.9.2. Hardy-Weinberg equilibrium analysis

For both the Xhosa and CMA cohort, all genotyped variants were tested for compliance with Hardy-Weinberg equilibrium (HWE) using Tools For Population Genetic Analyses v1.3 (TFPGA) (Miller, 1997). For variants for which less than five individuals were genotyped, an analogue to Fischer's Exact test was used to evaluate deviation from HWE. For variants where all genotypes contained more than five individuals, a Pearson's chi-square test (χ^2) was used. Significant deviation from HWE in a population was indicated by a *P*-value (*P*) of less than or equal to 0.01.

3.9.3. Testing variants for linkage disequilibrium

Investigating the patterns of linkage disequilibrium (LD) in the Xhosa and CMA cohorts is important, as they might differ from other African populations. Testing for the level of LD among variants enables the differentiation between variants that have a unique effect on EFV levels and treatment outcome, and linked variants with a mutual effect. Genotype data for all genotyped variants in the Xhosa and CMA cohorts that complied with HWE and had a MAF of greater than 0.01, were compiled and imported in Haploview v4.2 (Barrett *et al.*, 2005), using default settings. The parameters used to indicate LD among variants, were correlation coefficient (r^2) values of equal to or greater than 0.4 and logarithm of the odds (LOD) scores exceeding 2.

3.10. Statistical analysis

3.10.1. Analysing the effect of confounders

Given the small sample size of the study, adjustments for large numbers of confounding factors can reduce the power of statistical analysis for the study. Therefore it was decided to minimise

the number of adjustments. Using R statistical software (The R Development Core Team, 2010), the following confounding effects were tested for significant influence on the data generated: Ethnicity, time using EFV, baseline CD4, baseline VL, adherence, age at diagnosis, weight, smoking habits, alcohol consumption and WHO stage at treatment initiation. Confounding factors that showed significance ($P \leq 0.05$) were adjusted for in all subsequent statistical analysis.

3.10.2. Testing for correlations between genotype and EFV hair levels using linear regression models

Using Microsoft Office Excel 2010, the preliminary effect of genotype on mean EFV levels was investigated, before statistical adjustments were made to the data. The two separate Xhosa and CMA cohorts were sorted into groups according to genotype for each variant. From the number of patients in each genotype group, the number and frequencies of each allele was calculated. The mean EFV level was calculated for each group (genotype and allele) and compared.

Where possible, 2x2 and 3x2 contingency tables (www.vassarstats.net) were used to indicate significant differences in genotype frequency between groups. A *P*-value of ≤ 0.05 was an indication of possible significance. In each case where the individuals of one group were less than 5, a Fisher's exact test was used to test significance. In all other cases, a χ^2 test was used.

The linear effect of single variant genotype on EFV levels in the Xhosa and CMA cohorts were analysed using R statistical software (The R Development Core Team, 2010). The effects of statistically significant variants ($P \leq 0.05$) were quantified through linear regression models. All sequenced and genotyped variants that complied with HWE and had a MAF of ≥ 0.1 were included in the analysis.

Additionally, the effect of composite variant genotypes on EFV levels were also analysed, using genotype data of *CYP2B6*6* (*CYP2B6 516G>T*), as well as SNPs from *CYP1A2*, *CYP2A6*, *CYP3A4* and *CYP3A5* in the Xhosa and CMA cohort (Victor 2011; Röhrich 2012). *NR1I3* and *NR1I2* variants were analysed in combination with *CYP2B6 516G>T* in order to find out whether any of these variants either increase or diminish the elevated EFV levels caused by *CYP2B6 516G>T*. From previous results, it was found that in the current data, the *CYP2B6*6* allele has a recessive effect on EFV levels (Röhrich, 2012). Therefore, the two cohorts were separated into individuals that have two *CYP2B6*6*-alleles (indicated as *6*6), and individuals that have either no *CYP2B6*6* alleles or one *CYP2B6*6* allele (indicated as Non-*6*6). The *NR1I2* and *NR1I3* target SNPs were also analysed in conjunction with six SNPs from *CYP1A2* (*CYP1A2*1F*), *CYP2A6* (*CYP2A6*9B* and *CYP2A6*17*), *CYP3A4* (*CYP3A4*1B*) and *CYP3A5*

(*CYP3A5*6* and *CYP3A5*3*). Frequency data for these SNPs for the study cohort were available from a prior study (Victor, 2011). The empirical influence of statistically significant composite genotypes on EFV levels (effect size) were calculated using additive allelic models. The effect of haplotypes as identified through LD analysis on EFV levels was also analysed.

3.10.3. Testing for correlations between genotype and treatment outcome or adverse drug reactions using logistic regression models

Patient response to treatment was dichotomised, as indicated by VL measurements on the date of sample collection or the last measurement prior to sample collection. Patients with a VL of < 80 copies/ml were considered responders to treatment and patients with a VL of > 80 copies/ml were considered non-responders to treatment (Röhrich, 2012). The genotype and allele frequency for each variant for these two groups were compared, with a *P*-value of ≤ 0.05 as an indication of possible significance.

Patients were also dichotomised according to the presence or absence of EFV-based ADRs. The list of ADRs consisted of abnormal dreams, insomnia, depression and psychosis. Patients that had at least one clinical report of one or more of these ADRs were identified as "ADR-patients". Patients with an absence of these ADRs were identified as "non-ADR patients". The genotype and allele frequencies of each variant of these two groups were portrayed in column graphs.

The effect of genotype on two outcomes, response to treatment and risk of ADRs, also dichotomised as explained in Section 3.10.1.2, was analysed using R statistical software (The R Development Core Team, 2010). All sequenced and genotyped variants were investigated for significant effect on one or both outcomes. The effect of statistically significant variants ($P \leq 0.05$) was quantified through logistic regression models. All variants that complied with HWE and had a MAF of ≥ 0.1 were included in the analysis. Xhosa and CMA cohorts were analysed separately. The additive effect of identified haplotypes on the two outcomes, treatment outcome and the occurrence of ADRs, was also analysed.

CHAPTER 4

4. Results

4.1. Characteristics of study participants

A total of 136 female individuals were initially recruited, but only 132 participated in the current study. Participants were divided into two cohorts based on ethnicity as identified by the patients themselves. Participants that were excluded from the study included pregnant patients that were prescribed NVP instead of EFV (n = 1), according to the 2010 South African Guidelines (National Department of Health, 2010), patients providing insufficient DNA concentration (n = 1), patients with missing clinical folders (n = 2) and patients that were not of Xhosa or CMA ancestry (n = 1). A total of 79 patients were identified as Xhosa ethnicity and 53 patients were identified as Cape Mixed Ancestry (CMA) ethnicity. All patients were HIV-positive and prescribed an ARV regimen that specifically included the NNRTI efavirenz (EFV). All patients were recruited at the TC Newman HIV Clinic in Paarl in the Western Cape.

Table 9 presents the characteristics of the two cohorts used in this study. The Xhosa cohort had a median age of 39 years (inter-quartile range (IQR): 34 - 45) at the time of sample collection, which was slightly younger than the CMA cohort (median: 41 years, IQR: 37 - 46).

Baseline height and weight was used to calculate the baseline BMI for each patient. The median BMI for both cohorts fall between the range as described by the WHO as “normal” (World Health Organization, 2000), with the Xhosa median BMI (23.38, IQR: 21-30) slightly higher than the CMA median BMI (21.46, IQR: 19-24). Figure 8 shows the percentage of each cohort characterised as “underweight”, “normal” and “overweight” according to WHO criteria (World Health Organization, 2000), compared to the national South African population average as described in the WHO Global Database on Body Mass Index (World Health Organization, 2013c). Roughly half of the individuals in both cohorts have a normal BMI (Xhosa: 50.6%, CMA: 50.9%), which is slightly higher than the national South African percentage (46.2%) (World Health Organization, 2013c).

Time using EFV differed for the two cohorts. The majority of Xhosa individuals were taking an ARV regimen containing EFV (median: 41 months, IQR: 24 – 57 months) longer than the majority of CMA individuals (median: 26 months, IQR: 10 – 41 months).

Table 9: Demographic and clinical characteristics of study participants

	Xhosa	CMA
Characteristic	**Median (IQR)**	**Median (IQR)**
Age at sample collection (years)	39 (34 - 45)	41 (37 - 46)
BMI Baseline	23.39 (21.06 - 29.66)	21.46 (18.69 - 24.39)
CD4 Baseline (copies / ml)	159 (98 - 206)	147 (98 - 192)
Time using EFV (months)	41 (24 - 57)	26 (10 - 41)
Characteristic	**Number of pts (%)**	**Number of pts (%)**
WHO Disease Stage:		
Stage I	3 (3.80)	3 (5.66)
Stage II	16 (20.25)	12 (22.64)
Stage III	39 (49.37)	21 (39.62)
Stage IV	18 (22.78)	14 (26.42)
n/a	3 (3.80)	3 (5.66)
Smoking:		
yes	2 (2.53)	10 (18.87)
no	47 (59.49)	24 (45.28)
n/a	30 (37.97)	19 (35.85)
Alcohol:		
yes	5 (6.33)	9 (16.98)
no	48 (60.76)	26 (49.06)
n/a	26 (32.91)	18 (33.96)
Adherence		
good	70 (87.50)	43 (81.13)
acceptable	6 (7.50)	7 (13.21)
bad	3 (3.75)	2 (3.77)
n/a	1 (1.25)	1 (1.89)
Change in regimen		
No change	42 (53.16)	27 (50.94)
Change	37 (46.84)	26 (49.06)
More than once	9 (11.39)	7 (13.21)
Regimen during sampling		
EFV/3TC/d4T	46 (58.23)	33 (62.26)
EFV/3TC/AZT	21 (26.58)	9 (16.98)
EFV/3TC/TDF	12 (15.19)	10 (18.87)
n/a	0 (0.00)	1 (1.89)
Occurrence of ADRs		
yes	11 (13.92)	5 (9.43)
no	68 (86.08)	48 (90.57)
Second-line therapy		
yes	6 (7.59)	1 (1.89)
no	73 (92.41)	52 (98.11)
n/a	0 (0.00)	1 (1.89)
Additional medication		
yes	18 (22.78)	13 (24 53)
no	61 (77.22)	40 (75.47)

IQR, inter-quartile range. pts, patients. n/a, not applicable. BMI, body mass index. CD4, cluster of differentiation 4. EFV, efavirenz. 3TC, lamivudine. d4T, stavudine. AZT, zidovudine. TDF, tenofovir disoproxil fumarate.

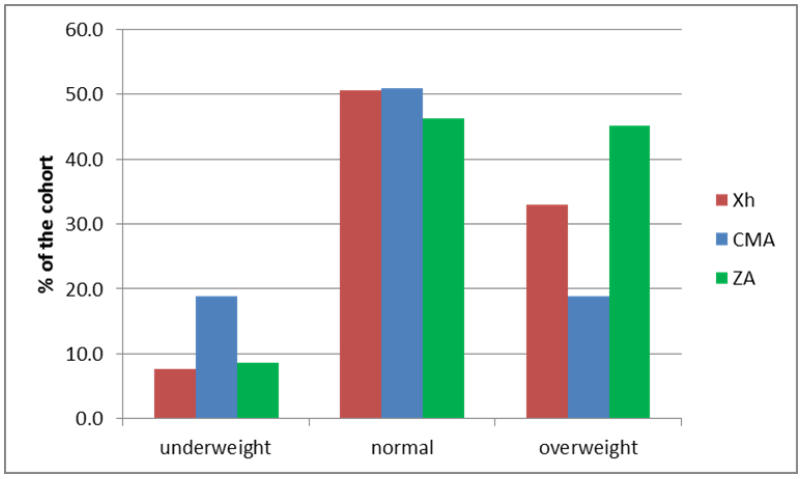


Figure 8: BMI distribution in the Xhosa and CMA cohorts compared to the nNational South African population as described in the WHO Global Database on Body Mass Index (World Health Organization, 2013c).
CMA, Cape Mixed Ancestry cohort; ZA, South African population.

4.2. *In silico* analysis of *NR1I3*

4.2.1. Analysis of sequence conservation using mVISTA

Sequence conservation in the human *NR1I3* gene was analysed through the comparison of the gene to eight orthologs from different species using mVISTA software. Figure 9 shows the percentage sequence identity in the 5000 bp upstream region, exon 1, intron 1 and exon 2 of each ortholog to the human *NR1I3* gene. Three important regions were conserved in all eight orthologs and were included in the regions used for bi-directional sequencing. The first region is a 260-bp region directly preceding the start of exon 1 and could play an important role in transcriptional regulation of the *NR1I3* gene. The other two conserved regions are located at exons 1 and 2; this shows the functional relevance of these areas as part of the coding region. Both these two conserved regions also include parts of intron 1, which could indicate an important role of these regions in the regulation of intron splicing.

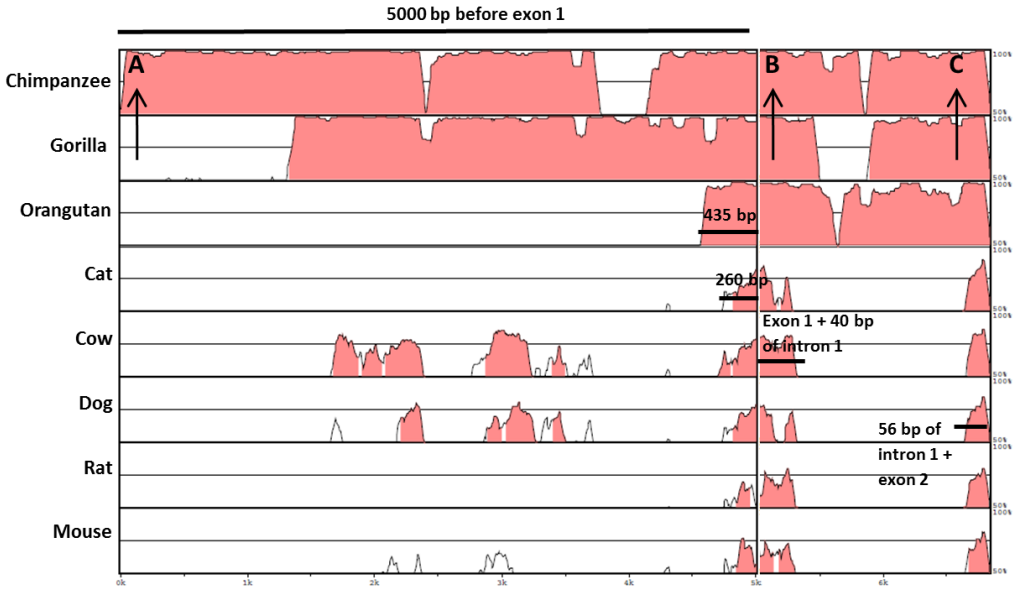


Figure 9: mVISTA sequence alignment of eight orthologs to the human *NR1I3* gene.
The pink areas indicate the % identity of the ortholog sequence to the human *NR1I3* gene. The sequence starts 5000 kb before exon 1 and extends until the end of exon 2. The arrows indicate the location of functionally identified TFBSs on the human *NR1I3* gene. Arrow A indicates a glucocorticoid receptor binding site as identified by Pascussi *et al*. 2003. Arrow B indicates an HNF4α2 binding site as identified by Pascussi *et al*. 2007. Arrow C indicates a serum response element binding site as identified by Osabe *et al*. 2009.

4.2.2. Identification of putative transcription factor binding sites using freeware prediction programs

Six prediction programs were used to find putative TFBSs in the 5 kb upstream region of *NR1I3*, as well as Exon 1 and Intron 1. All putative TFBSs detected in at least three of the six programs (n = 17) are listed in Table 10, as well as the three functionally identified TFBS from literature. Two of the three functional TFBSs, a glucocorticoid receptor (GR) and a hepatic nuclear factor 4α (HNF-4α), were also detected by three or more prediction programmes. The third functional TFBS, a serum response element (SRE) was only detected in rVISTA. A 500-bp region located 4.6 kb from the start of exon 1 contained five of the 17 identified putative TFBS, and included the functional GR. The high occurrence of putative TFBSs prompted the addition of this region into the list of amplicons used for bi-directional sequencing. Figure 9 shows the location of three functionally identified TFBS (arrows A, B and C). Only one TFBS, an HNF4α2 binding site (arrow B), is located in one of the conserved regions of the *NR1I3* gene.

Table 10: Putative and functional TFBS in *NR1I3* as identified in TFBS prediction software and literature.

Position (bp) from start of exon 1	TF	TFBS prediction software						Literature
		Match	AliBaba	Patch	P-Match	Jaspar	rVista	
-4683	GR		x	x			x	Pascussi *et al*. 2003
-4541	AP-1	x	x	x		x	x	n/a
-4492	GATA-3	x	x	x		x	x	n/a
-4429	GATA-3	x	x	x		x	x	n/a
-4201	AP-1	x	x	x		x	x	n/a
-3392	C/EBP	x	x		x	x	x	n/a
-2517	C/EBP	x	x	x		x	x	n/a
-1890	C/EBP	x	x		x	x	x	n/a
-1651	C/EBP	x		x	x	x	x	n/a
-1186	USF	x	x	x		x	x	n/a
-260	C/EBP	x		x	x	x	x	n/a
-61	AP-1	x	x		x	x	x	n/a
+25	HNF-4α	x		x		x	x	Pascussi *et al*. 2007
+305	GATA-3	x	x		x	x	x	n/a
+337	USF	x	x		x	x	x	n/a
+411	C/EBP	x		x	x	x	x	n/a
+622	C/EBP	x		x	x	x	x	n/a
+1611	SRF						x	Osabe *et al*. 2009

The symbol "x" signifies that this TFBS was predicted in the corresponding program. GR, glucocorticoid receptor. AP-1, activator protein 1. GATA-3, GATA-binding protein 3. C/EBP, CCAAT/enhancer-binding protein. USF, upstream stimulatory factor. HNF-4, hepatic nuclear factor 4. SRF, serum response factor.

4.3. *NR1I3* variants identified through bi-directional sequencing

Bi-directional sequencing results of all amplicons of the 15 Xhosa and 15 CMA individuals were aligned to a reference sequence from the Ensembl Genomes Browser, and single nucleotide differences that occurred in both sequence directions were identified as variants. The chromatograms for all identified variants are illustrated in Appendix F.

4.3.1. Incidence of short tandem repeats (STRs)

Sequencing results for Amplicons UA2, UA3 and Exon 6/7 showed the occurrence of enzyme slippage due to the presence of short tandem repeats (STRs) (Figure 10). Subsequently, an alternative forward primer was designed for Exon 6/7 in order to avoid the STR region. The design of alternative primer sets for UA2 and UA3 was not viable, due to a high number of STR regions and poly T regions in these amplicons (see Appendix E). Therefore it was decided to exclude these two amplicons from the region used for bi-directional sequencing.

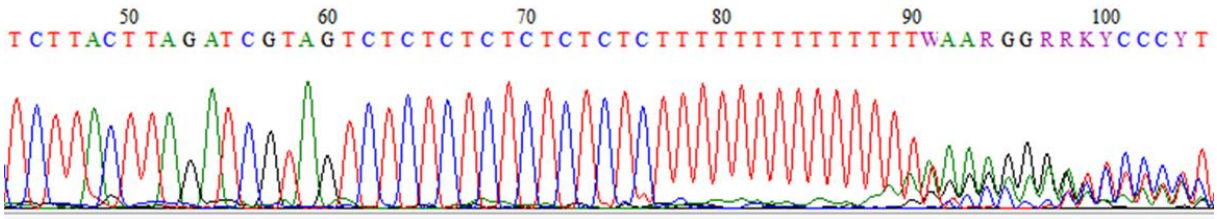


Figure 10: Chromatogram showing the occurrence of enzyme slippage in amplicon UA3 due to a TC-allele repeat followed by a poly-T-allele repeat.

4.3.2. Identification of target *NR1I3* variants for genotyping

Following alignment of all sequenced amplicons to the reference sequence ENST00000367983 from the Ensembl Genomes Browser, a total of 12 SNPs were identified, three 5'-upstream SNPs, six intronic SNPs and three exonic synonymous SNPs. Table 11 lists these 12 SNPs and their MAFs in the sequenced Xhosa and CMA cohorts, as well as in the Yoruban population from the 1000 Genomes Study, which contributes partially to the genetic constitution of the Xhosa population and, to a lesser extent, to the CMA population (Tishkoff *et al*., 2009). As mentioned previously, the DBD encoding region and parts of the LBD encoding region of *NR1I3* was included in the sequenced amplicons (Section 3.6.1). Only one SNP was identified in the DBD region of *NR1I3*. The LBD region of *NR1I3* is 3,458 bp long, however, because the LBD region also includes introns, only fragments amounting to a combined length of 1,646 bp were included in the sequencing amplicons. Five of the 12 identified SNPs fall within these fragments.

From the 12 SNPs identified from sequencing, only six were selected for genotyping in the entire patient cohort. Three SNPs (rs55802895, rs2502815, rs2307424) were previously reported to have functional importance (Urano *et al*., 2009; Ikeda *et al*, 2010; Wyen *et al*., 2011; Swart *et al*., 2012), of which one SNP (rs2307424) showed specific importance to EFV treatment outcome (Wyen *et al*., 2011; Swart *et al*., 2012). One of these SNPs (rs55802895) also disrupts a potential retinoic acid receptor-alpha (RAR-α) binding site, as predicted in AliBaba (Grabe, 2002). One SNP (rs35205211) was selected because it disrupts potential mRNA folding patterns, as predicted in mFOLD (Zuker, 2003). From the pool of intronic SNPs that showed no further importance in literature, SNPs rs35709078 and rs2307420 were also selected for genotyping due to their low MAF in the YRI population, which could indicate that they would also be present in the Xhosa and CMA cohorts.

Three additional SNPs that were not identified in sequencing results were selected for genotyping based on their functional importance in literature and are indicated in bold in Table 11. The first of these SNPs, rs113800307 (Met265Val), was selected because of its identification in a whole-

genome sequencing study using South African Khoisan and Bantu individuals (Schuster *et al*., 2010). The second SNP, rs34161743 (R97W), potentially alters protein activity as predicted by both SIFT and PolyPhen, (Ramensky *et al*., 2002; Ng, 2003) and disrupts a potential ESE (Thompson *et al*. 2005). Lastly, rs3003596, an intronic SNP, was selected for genotyping as this SNP was previously significantly associated with decreased EFV plasma levels (Swart *et al*., 2012).

4.4. *NR1I2* variants identified through a survey of literature and genotype databases

From data obtained from previous studies and from the HapMap and 1000 Genomes database as described in Section 3.7 in Chapter 3, a total of ten variants were selected for genotyping, three 5'-upstream variants (rs1523130, rs3814055, rs1523128), two 5'-UTR variants (rs1523127, rs38426891), four intronic variants (rs2472677, rs1464603, rs1464602, rs2276707) and one non-synonymous variant (rs12721613). These identified variants are listed in Table 12. As no previous frequency data are available for these *NR1I2* variants for the Xhosa and CMA cohorts, the MAF data for four other ethnic groups (CEU, CHB, YRI and ASW, respectively representing Caucasian, Asian, African and African American populations), obtained from the 1000 Genomes database and from literature, are displayed in Table 12 as guidelines. Based on the genetic contribution of each of the four populations to the Xhosa and CMA cohorts, frequency data for the Xhosa cohort are expected to be most similar to the Yoruban frequency data, and for the CMA cohort are expected to present as intermediate between either the Caucasian and Yoruban frequency data, or the Asian and Yoruban frequency data (Tishkoff *et al*., 2009). All variants except for two (rs1523130 and rs1464603) have MAFs greater than 0.05 in the YRI population, based on 1000 Genomes data. These two SNPs, however, have high MAFs in the other three populations (CEU, CHB and ASW), which could suggest that these SNPs could have a high MAF in the CMA population.

The functional importance of the identified variants is indicated in Table 12. The six variants located in the 5'-upstream, 5'-UTR and Intron 1 region all disrupt putative TFBSs (Zhang *et al*., 2001; Lamba *et al*., 2008). Five variants play a role in the regulation of *CYP3A4*. SNPs rs1523130 and rs1523127 are significantly associated with CYP3A4 expression (King *et al*., 2007; Lamba *et al*., 2008); SNPs rs3814055, rs2472677 and rs12721613 are significantly associated with *CYP3A4* induction (Lamba *et al*., 2008; Svärd *et al*., 2010); and rs1523130 and rs2472677 are significantly associated with *CYP3A4* activity (Lamba *et al*., 2008).

Table 11: *NR1I3* variants identified from bi-directional sequencing and literature, and their MAFs in the Xhosa, CMA and YRI populations

Location	rs number	Genomic identifier	Amino acid change	Xhosa		CMA		YRI	Putative TFBS (AliBaba)	Functional Importance
				n	MAF	n	MAF	MAF		
5'-upstream	rs55802895	258G>A	n/a	3	0.100	3	0.100	0.034	RAR-α	↑ risk of hyperlipidemia[1]
5'-upstream	rs2501870	432C>T	n/a	7	0.267	8	0.267	0.426	n/a	n/a
5'-upstream	rs11584174	548G>A	n/a	5	0.200	4	0.133	0.159	n/a	n/a
Intron 2	rs35709078	7145C>T	n/a	1	0.033	1	0.033	0.023	n/a	n/a
Intron 3	rs141260451	7393C>G	n/a	0	0.000	1	0.033	0.000	n/a	n/a
Intron 3	**rs3003596**	**8784T>C**	n/a	**n/a***	**n/a***	**n/a***	**n/a***	**0.386**	n/a	**↓ EFV plasma levels[2]**
Intron 3	rs2502815	9774C>T	n/a	5	0.233	9	0.333	0.420	n/a	↓ bone mineral density[3]
Exon 4	rs35205211	9892C>G	n/a	1	0.033	1	0.033	0.006	n/a	Disrupts putative ssDNA and mRNA folding patterns (mFOLD)
Exon 4	**rs34161743**	**9923C>T**	R97W	**0**	**0.000**	**0**	**0.000**	**0.017**	n/a	**Deleterious and probably damaging amino acid change (SIFT, PolyPhen)**
Exon 5	rs2307424	10396C>T	n/a	1	0.033	3	0.100	0.108	n/a	↓ EFV discontinuation[2,4], ↓ EFV plasma levels[2], ↑ risk of hyperlipidemia[1]
Intron 5	rs2307420	10445A>G	n/a	4	0.133	3	0.100	0.028	n/a	n/a
Exon 7	**rs113800307**	**12076A>G**	**M265V**	**0**	**0.000**	**0**	**0.000**	**n/a**	n/a	**Identified in Bantu (Xhosa/Venda) individuals[5]**
Intron 8	rs34859188	13196T>C	n/a	3	0.133	3	0.100	0.023	n/a	n/a
Intron 8	rs35103872	13220G>A		3	0.100	3	0.100	0.017	n/a	n/a
Exon 9	rs36017137	13297G>A		1	0.033	0	0.000	0.006	n/a	n/a

SNP frequencies are displayed for the following groups: The Xhosa and Cape Mixed Ancestry cohorts of this study and the Yoruban population from the 1000 Genomes Project. The shaded regions are the SNPs selected from sequencing results for genotyping. The SNPs in bold are additional SNPs selected from literature for genotyping. ↑, increased. ↓, decreased. * No available Xhosa and CMA frequency data, as SNPs are outside the regions of *NR1I3* selected for sequencing. n/a, not available. [1]Ikeda *et al*., 2010. [2]Swart *et al*., 2012. [3]Urano *et al*., 2009. [4]Wyen *et al*., 2011. [5]Schuster *et al*., 2010.

Table 12: *NR1I2* variants selected for genotyping, their MAFs in four population groups from the 1000 Genomes Project and their functional importance.

Location	rs number	Genomic identifier	Amno acid change	MAF				Putative TFBS[1,3]	Functional importance
				Cau (CEU)	Asn (CHB)	Afr (YRI)	AA (ASW)		
5' upstream	rs1523130	5177T>C	n/a	0.359 (T)	0.268 (T)	0.04 (C)	0.139 (C)	STAT1,3,6 and NFAT	↓ basal *CYP3A4* activity, ↓ *CYP3A4* and PXR expression[1]
5' upstream	rs3814055	5705C>T	n/a	0.341	0.268	0.301	0.238	NF-KB and ISGF-3	↓ PXR expression, ↓ *CYP3A4* induction[1]; ↑ cyclosporine bioavailabilty[4]
5' upstream	rs1523128	6334A>G	n/a	0.000	0.000	0.432	0.270	C/EBP and HNF-1α	In LD with rs3814055, rs1523127 and rs2276707, ↑ susceptibility to Crohn's Disease[8]
5' UTR	rs1523127	6709C>A	n/a	0.341 (C)	0.268 (C)	0.097 (A)	0.230 (A)	n/a	↓ *ABCB1* and *CYP3A4* expression[5]
5'UTR	rs3842689[1]	7070_7075del	n/a	0.290[1]	n/a	0.173[2]	n/a	HNF-1α	↓ PAR2 expression[1]; ↑ NVP clearance[2]
Intron 1	rs2472677	24087C>T	n/a	0.359 (C)	0.371 (C)	0.352 (T)	0.361 (T)	HNF3β	↑basal *CYP3A4* activity, ↓ *CYP3A4* induction, ↑ PAR2 expression, ↑ PXR transcription[1]; ↓ plasma atazanavir levels[6]
Exon 2	rs12721613	31846C>T	P27S	0.000	0.000	0.188	0.107	n/a	n/a
Intron 2	rs1464603	32019G>A	n/a	0.318 (G)	0.345 (G)	0.023 (A)	0.197 (A)	n/a	In LD with rs1464602 and rs3732357, ↓ plasma midazolam clearance[7]
Intron 2	rs1464602	32042G>A	n/a	0.318 (G)	0.345 (G)	0.233 (A)	0.369 (A)	n/a	In LD with rs 1464603 and rs3732357, ↓ plasma midazolam clearance[7]
Intron 6	rs2276707	39823C>T	n/a	0.118	0.459	0.386	0.352	n/a	In LD with rs3814055, rs1523127, rs3842689 and rs6785049, ↑ cyclosporine bioavailability[4]

CEU, CHB, YRI and ASW, represent four respective ethnic groups from the 1000 Genomes Project. CEU, Residents in Utah, USA, of Northern and Western European ancestry. CHB, Han Chinese in Bejing, China. YRI, Yoruba in Ibadan, Nigeria. ASW, Americans of African Ancestry in South West, USA. n/a, not available. ↑, increased. ↓, decreased. [1]Lamba *et al*. 2008. [2]Brown *et al*. 2012, [3]Zhang *et al*., 2001, [4] Fanta *et al*., 2010, [5] King *et al*., 2007, Siccardi *et al*., 2008, [7] He *et al*., 2006. [8]Glas *et al*., 2011.

Six variants play a role in the regulation of *NR1I2*. SNPs rs1523130 and rs2472677 are associated with *NR1I2* transcription (Lamba *et al*., 2008); and SNPs rs1523130, rs3814055 and rs2472677 and deletion rs3842689 are associated with *NR1I2* expression (Lamba *et al*., 2008). Five variants play a role in the pharmacokinetic characteristics of certain drugs. SNPs rs3814055 and rs2276707 are significantly associated with cyclosporine bioavailability (Fanta *et al*., 2010); deletion rs3842689 is significantly associated with nevirapine clearance (Brown *et al*., 2012); and SNPs rs1464603 and rs1464602 are significantly associated with midazolam clearance (He *et al*., 2006).

4.5. Variant genotyping results

Following the selection of SNPs to be genotyped (Section 4.3.2 and 4.4), a total of 19 variants from *NR1I3* and *NR1I2* were genotyped for both cohorts (Xhosa n = 79 and CMA n = 53). Appendix G shows examples of gel photos for TSP, RFLP and deletion assays analysis of each variant for heterozygous individuals, individuals homozygous for the major allele and individuals homozygous for the minor allele, where applicable. Appendix H shows the allelic discrimination plots for Taqman analysis of rs1523130 (5177T>C) for all samples, and also amplification plots for one patient that is homozygous for the major allele (TT), one patient that is heterozygous (TC) and one patient that is homozygous for the minor allele (CC).

4.5.1. Analysis of genotype and allele frequencies

All variants genotyped were polymorphic in both the Xhosa and CMA cohorts, except for rs113800307, which was non-polymorphic in both cohorts. Table 13 lists the observed genotype frequencies and MAFs of all variants in the Xhosa and CMA cohorts.

Four out of the 19 variants (21.05%), indicated in bold in Table 13, differed significantly between the cohorts with regards to allele frequency (a P-value of lower than or equal to 0.05). Of these four variants, one was located in *NR1I3* (rs2307424) and three were located in *NR1I2* (rs1523130, rs3814055, rs1523127). The MAF for rs3814055 was higher in the Xhosa cohort than in the CMA cohort, while the MAFs for rs2307424, rs1523130 and rs1523127 were higher in the CMA cohort.

Table 13: Genotype frequencies and MAFs for *NR1I3* and *NR1I2* variants in the Xhosa and CMA cohorts

Gene	Location	rs name	Genomic identifier	Genotype	Xhosa n = 79 [1]		Cape Mixed Ancestry n = 53 [1]		*P*-value[3]
					n (frequency)	MAF	n (frequency)	MAF	
				GG	67 (0.848)		46 (0.868)		
NR1I3	5'-upstream	rs55802895	258G>A	GA	11 (0.139)	0.082	7 (0.132)	0.066	0.624
				AA	1 (0.013)		0 (0.000)		
				CC	76 (0.962)		52 (0.981)		
NR1I3	Intron 2	rs35709078	7145C>T	CT	3 (0.038)	0.019	1 (0.019)	0.009	0.651
				TT	0 (0.000)		0 (0.000)		
				TT	35 (0.449)		18 (0.340)		
NR1I3	Intron 3	rs3003596	8784T>C	TC	33 (0.423)	0.340	26 (0.491)	0.415	0.215
				CC	10 (0.128)		9 (0.170)		
				CC	43 (0.551)		24 (0.462)		
NR1I3	Intron 3	rs2502815	9774C>T	CT	28 (0.359)	0.269	23 (0.442)	0.317	0.403
				TT	7 (0.090)		5 (0.096)		
				CC	74 (0.974)		51 (0.981)		
NR1I3	Exon 4	rs35205211	9892C>G	CG	2 (0.026)	0.013	1 (0.019)	0.010	1.000
				GG	0 (0.000)		0 (0.000)		
				CC	76 (0.962)		51 (0.981)		
NR1I3	Exon 4	rs34161743	9923C>T	CT	3 (0.038)	0.019	1 (0.019)	0.010	0.654
				TT	0 (0.000)		0 (0.000)		
				CC	**67 (0.848)**		**37 (0.698)**		
NR1I3	**Exon 5**	**rs2307424**	**10396C>T**	**CT**	**12 (0.152)**	**0.076**	**13 (0.245)**	**0.179**	**0.011**
				TT	**0 (0.000)**		**3 (0.057)**		
				AA	70 (0.886)		47 (0.887)		
NR1I3	Intron 5	rs2307420	10445A>G	AG	9 (0.114)	0.057	5 (0.094)	0.066	0.796
				GG	0 (0.000)		1 (0.019)		
				AA	78 (1.000)		41 (1.000)		
NR1I3	Exon 7	rs113800307	12076A>G	AG	0 (0.000)	0.000	0 (0.000)	0.000	1.000
				GG	0 (0.000)		0 (0.000)		
				TT	**61 (0.782)**		**31 (0.596)**		
NR1I2	**5'-upstream**	**rs1523130**	**5177T>C**	**TC**	**14 (0.179)**	**0.128**	**14 (0.269)**	**0.269**	**0.004**
				CC	**3 (0.038)**		**7 (0.135)**		
				CC	**48 (0.640)**		**42 (0.824)**		
NR1I2	**5'-upstream**	**rs3814055**	**5705C>T**	**CT**	**21 (0.280)**	**0.220**	**9 (0.176)**	**0.088**	**0.018**
				TT	**6 (0.080)**		**0 (0.000)**		
				GG	41 (0.526)		33 (0.623)		
NR1I2	5'-upstream	rs1523128	6334A>G	AG	32 (0.410)	0.269	18 (0.340)	0.208	0.254
				AA	5 (0.064)		2 (0.038)		
				CC	**50 (0.649)**		**23 (0.442)**		
NR1I2	**5'-UTR**	**rs1523127**	**6709C>A**	**CA**	**23 (0.299)**	**0.201**	**20 (0.385)**	**0.365**	**0.003**
				AA	**4 (0.052)**		**9 (0.173)**		
				+/+	50 (0.641)		41 (0.774)		
NR1I2	5'-UTR	rs3842689[2]	7070_7075del GAGAAG	+/-	23 (0.295)	0.212	10 (0.189)	0.132	0.100
				-/-	5 (0.064)		2 (0.038)		
				CC	30 (0.385)		18 (0.360)		
NR1I2	Intron 1	rs2472677	24087C>T	CT	35 (0.449)	0.391	27 (0.540)	0.370	0.806
				TT	13 (0.167)		5 (0.100)		

Gene	Location	rs name	Genomic identifier	Genotype	Xhosa n = 79 [1]		Cape Mixed Ancestry n = 53 [1]		*P*-value[3]
					n (frequency)	MAF	n (frequency)	MAF	
NR1I2	Exon 2	rs12721613	31846C>T	CC	63 (0.808)	0.096	49 (0.925)	0.038	0.091
				CT	15 (0.192)		4 (0.075)		
				TT	0 (0.000)		0 (0.000)		
NR1I2	Intron 2	rs1464603	32019G>A	GG	51 (0.646)	0.190	32 (0.604)	0.226	0.471
				GA	26 (0.329)		18 (0.340)		
				AA	2 (0.025)		3 (0.057)		
NR1I2	Intron 2	rs1464602	32042G>A	GG	36 (0.456)	0.329	21 (0.412)	0.402	0.232
				GA	34 (0.430)		19 (0.373)		
				AA	9 (0.114)		11 (0.216)		
NR1I2	Intron 6	rs2276707	39823C>T	CC	23 (0.295)	0.449	23 (0.442)	0.365	0.182
				CT	40 (0.513)		20 (0.385)		
				TT	15 (0.192)		9 (0.173)		

Variants that differed significantly between the Xhosa and CMA cohort are indicated in bold. [1]Due to problems with amplification, for some variants, the genotypes for some samples in the cohorts are not available. The true number of samples genotyped for each variant is stated in columns 6 and 8 for each separate genotype. [2]The genotypes of the 6 bp-deletion are described as follows: "+", one copy of the GAGAAG region is present; "-", one copy of the GAGAAG region is deleted. [3]A *P*-value of ≤ 0.05 is considered a significant difference in MAF.

The MAFs of all variants in both the Xhosa and CMA cohorts were compared to the MAFs in five populations from the 1000 Genomes Project and one population from the HapMap Project. Population data was unavailable for SNP rs113800307 and deletion rs3842689. Frequency data for the HapMap GIH population were unavailable for nine of the variants. Figure 11 shows the MAFs for the abovementioned four SNPs (rs2307424, rs1523130, rs3814055 and rs1523127) in these six populations, as well as in the study cohorts.

The range of variability in variant MAF among populations is demonstrated in Table 14 by the number of target SNPs that significantly differs in MAF between each population and the two genotyped cohorts ($P \leq 0.05$). The CEU, CHB and GIH populations showed the highest percentage of SNPs that differed significantly from the Xhosa and CMA cohorts (Table 14). The number of SNPs in the CEU population that differed in MAF from the Xhosa cohort was 13 out of 17 (76.47%) and from the CMA cohort was 12 out of 17 (70.59%). The number of SNPs that differed between the CHB population and the Xhosa and CMA cohorts with regards to MAF was 12 out of 17 (70.59). The percentage of SNPs that differed in MAF between the CIH population and the Xhosa and CMA cohorts cohorts was also high (75.00% and 63.50% respectively). The LWK and ASW populations showed the lowest percentage of SNPs that differed significantly from the Xhosa and CMA cohorts. The number of SNPs in the LWK population that differed in MAF from the Xhosa cohort was three out of 17 (17.65%), which is lower than the number for

the CMA cohort (eight out of 17, 47.06%). The number of SNPs in the ASW population that differed in MAF from the Xhosa cohort was three out of 17 (17.65%) and from the CMA cohort was four out of 17 (23.53%). Nine out of the total of 17 SNPs differed significantly in terms of MAF between the YRI population and both study cohorts (52.94%).

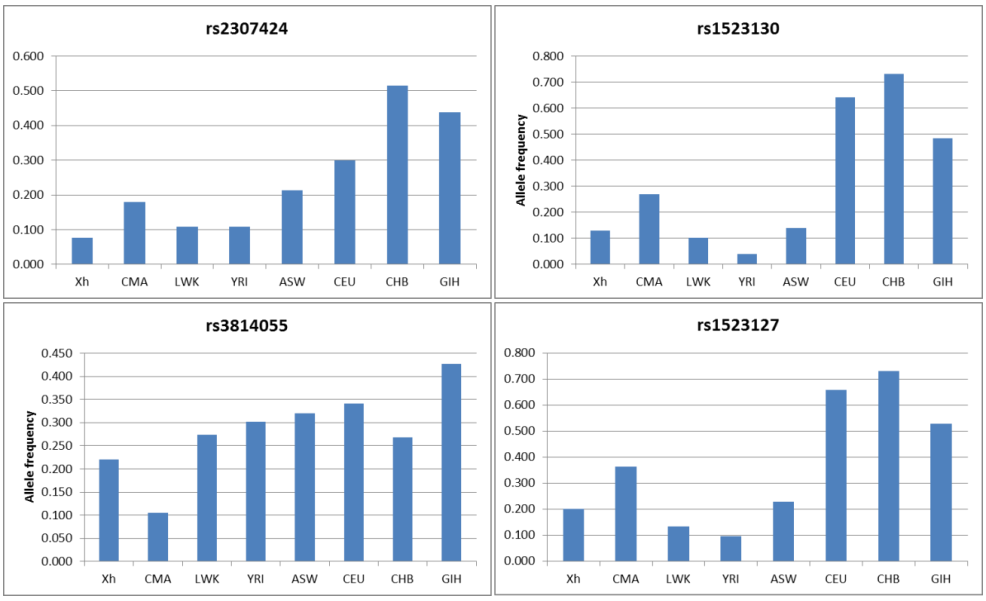


Figure 11: MAFs for rs2307424, rs1523130, rs3814055 and rs1523127 in the study cohorts and populations from the 1000 Genomes and HapMap databases.
CMA, Cape Mixed Ancestry cohort. LWK, Luhya in Webuye, Kenya. YRI, Yoruba in Ibadan, Nigeria. ASW, Americans of African Ancestry in South West, USA. CEU, Residents in Utah, USA, of Northern and Western European ancestry. CHB, Han Chinese in Bejing, China. GIH, Gujarati Indian from Houston, Texas.

Table 14: Number of SNPs in six populations that differed significantly in MAF from the Xhosa and CMA cohorts

Population (total)	Xhosa		CMA	
	n	%	n	%
LWK (17)	3	17.65	8	47.06
YRI (17)	9	52.94	9	52.94
ASW (17)	3	17.65	4	23.53
CEU (17)	13	76.47	12	70.59
CHB (17)	12	70.59	12	70.59
GIH (8)	6	75.00	5	62.50

The numbers in brackets are the total number of SNPs for which frequency data are available. A *P*-value of ≤ 0.05 is considered a significant difference in MAF. LWK, Luhya in Webuye, Kenya. YRI, Yoruba in Ibadan, Nigeria. ASW, Americans of African Ancestry in South West, USA. CEU, residents in Utah, USA, of Northern and Western European ancestry. CHB, Han Chinese in Bejing, China. GIH, Gujarati Indians from Houston, Texas

4.5.2. Results of Hardy-Weinberg equilibrium analysis

Eight variants in the Xhosa cohort and seven variants in the CMA cohort were tested for HWE using a χ^2 test; the remaining variants were tested using an analogu to Fisher's Exact test. In both the Xhosa and CMA cohorts, all variants fulfilled the criteria for HWE, therefore genotyping error and evolutionary forces could be ignored.

4.5.3. Linkage disequilibrium analysis of variants

Using Haploview v4.2 (Barrett *et al*., 2005), linkage disequilibrium analysis was performed for seven *NR1I3* variants and nine *NR1I2* variants. Variants that did not qualify for linkage disequilibrium analysis were rs113800307, which was non-polymorphic in both cohorts and rs35709078, which had a MAF of <0.01 in the CMA population. LD plots showed that no LD was found between any of the *NR1I3* variants in both Xhosa and CMA cohorts (Appendix I). LD plots for *NR1I2* variants in both cohorts are shown in Figure 12. In both cohorts, rs3814055 (5705C>T) was found to be in LD with the deletion rs3842689 (r^2 = 0.885 and 0.907 for Xhosa and CMA respectively; LOD score = 21.670 and 11.040 for Xhosa and CMA respectively) and rs1523130 (5177T>C) was found to be in LD with rs1523127 (6709C>A) (r^2 = 0.411 and 0.587 for Xhosa and CMA respectively; LOD score = 8.04 and 10.83 for Xhosa and CMA respectively).

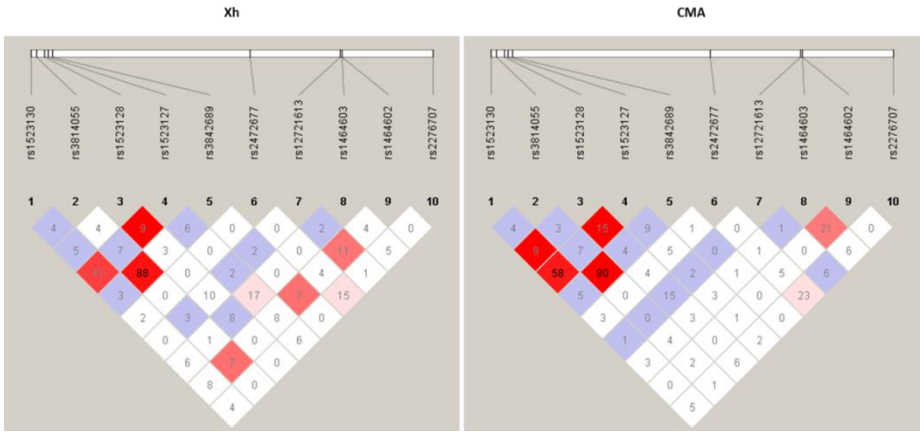


Figure 12: LD plots for *NR1I2* variants in the Xhosa and CMA cohorts.
LD values are represented in the blocks by correlation coefficient (r^2) values.

4.6. Analysing the effect of confounders

All possible confounding factors were tested for significant effect on EFV hair levels, treatment outcome and occurrence of ADRs. The three confounding factors that had a significant effect on

the three outcomes were time using EFV, age at diagnosis and ethnicity. Subsequently, statistical analysis was adjusted for these three confounders.

4.7. Correlations between genotype and EFV hair levels

4.7.1. The effect of single variant genotype on EFV levels

The preliminary effect of single variant genotype on EFV levels was analysed, in order to see if there are any trends that point to possible findings from the more detailed statistical analysis to follow. Table 15 shows the mean EFV levels for the genotype groups for each variant, separated by ethnicity.

Figure 13 shows the mean EFV levels for variants rs3842689 (7070_7075delGAGAAG), rs35205211 (9892C>G), rs1523130 (5177T>C) and rs1523128 (6334A>G). The mean EFV levels for all variants are shown in Appendix J. The highest mean EFV levels (12.76 ng/ml) were recorded for CMA individuals homozygous for the deletion rs3842689 (7070_7075delGAGAAG) (n = 2) (Table 15). The lowest mean EFV levels (1.15 ng/ml) were recorded for a CMA individual heterozygous for rs35205211 (9892C>G) (Table 15). There were no individuals homozygous for rs35205211 in the CMA cohort. For SNP rs1523130, individuals homozygous for the minor allele (CC) (n = 3) have the highest mean EFV level among the Xhosa cohort (10.42 ng/ml). Xhosa individuals homozygous for the minor alleles of rs1523128 and rs3814055 had the lowest EFV levels (3.53 ng/ml and 3.56 ng/ml respectively) among the Xhosa cohort, however, these levels were not as low as the lowest EFV levels for the CMA cohort.

Table 15: Genotype mean EFV levels for each variant for the Xhosa and CMA cohort

Gene	rs number	Genomic identification	Genotype	Mean EFV hair levels	
				Xhosa	CMA
NR1I3	rs2307420	10445A>G	AA	5.96	6.28
			AG	9.28	5.37
			GG	n/a	5.07
NR1I3	rs2307424	10396C>T	CC	6.11	6.00
			CT	7.66	6.91
			TT	n/a	5.14
NR1I3	rs34161743	9923C>T	CC	6.34	6.14
			CT	6.36	10.00
			TT	n/a	n/a
NR1I3	rs35205211	9892C>G	CC	6.31	6.32
			CG	8.15	1.15
			GG	n/a	n/a
NR1I3	rs2502815	9774C>T	CC	6.27	6.20
			CT	6.66	5.64
			TT	5.57	5.60
NR1I3	rs35709078	7145C>T	CC	6.37	6.22
			CT	5.60	3.98
			TT	n/a	n/a
NR1I3	rs55802895	258G>A	GG	6.43	6.27
			GA	5.92	5.53
			AA	5.37	n/a
NR1I2	rs1523130	5177T>C	TT	5.84	6.69
			TC	7.37	5.59
			CC	10.42	4.57
NR1I2	rs3814055	5705C>T	CC	6.61	5.51
			CT	6.78	7.81
			TT	3.56	4.62
NR1I2	rs1523128	6334A>G	GG	5.64	6.37
			GA	7.66	5.95
			AA	3.53	4.93
NR1I2	rs1523127	6709C>A	CC	5.83	6.41
			CA	4.93	6.43
			AA	7.66	4.97
NR1I2	rs3842689	7070_7075del GAGAAG	+/+	6.47	5.39
			+/-	6.58	8.08
			-/-	3.89	12.76
NR1I2	rs2472677	24087C>T	CC	5.91	6.92
			CT	7.00	5.86
			TT	5.52	6.14
NR1I2	rs12721613	31846C>T	CC	5.87	6.22
			CT	8.32	5.61
			TT	n/a	n/a
NR1I2	rs1464603	32019G>A	GG	5.93	6.80
			GA	7.32	5.09
			AA	4.08	6.04
NR1I2	rs1464602	32042G>A	GG	5.22	5.58
			GA	7.05	7.40
			AA	8.14	5.55
NR1I2	rs2276707	39823C>T	CC	6.90	6.06
			CT	6.43	5.09
			TT	5.22	8.93

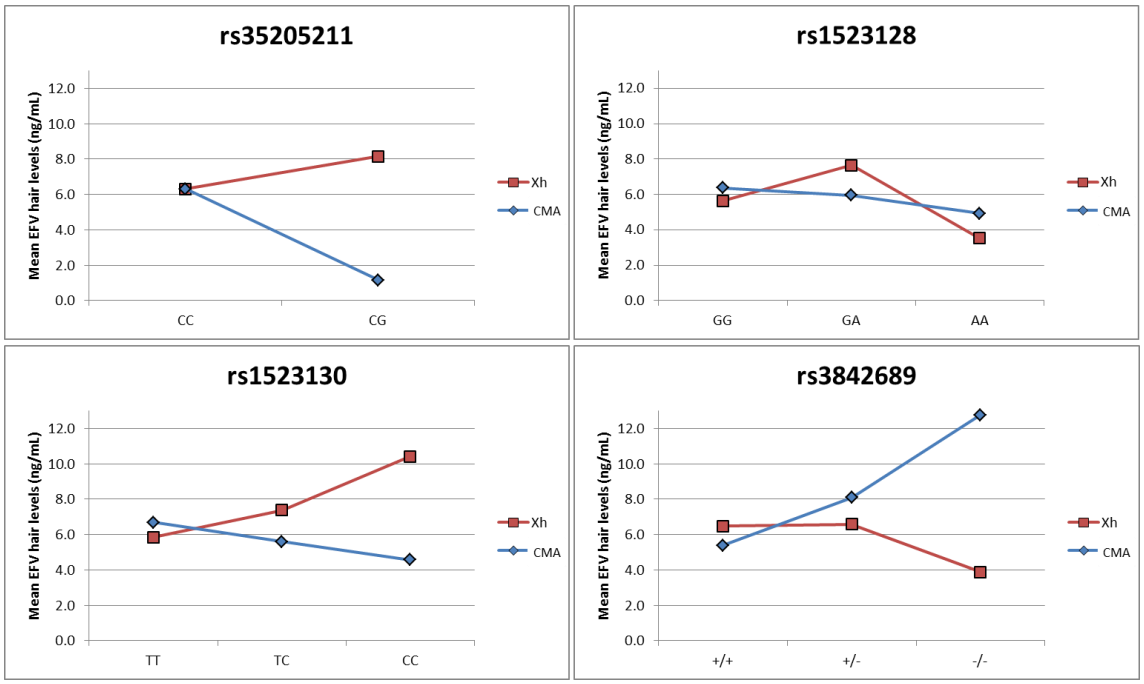


Figure 13: Mean EFV levels for rs35205211, rs1523130, rs1523128 and rs3842689 according to genotype.

The effect of single *NR1I3* and *NR1I2* variant genotypes on EFV was analysed using linear regression models in R statistical software (The R Development Core Team, 2010). In both cohorts, the natural logarithm of EFV levels was used, as the distribution of EFV levels is positively skewed. The *P*-values obtained following linear regression analysis for all variants are shown in Table 16. For single variant analysis, one *NR1I2* SNP, rs1523128 (6334A>G), was significantly associated with EFV levels ($P = 0.0168$), when correcting for time using EFV, age at diagnosis and ethnicity. In all individuals homozygous for the minor allele of rs1523128 (AA), EFV levels were 56% lower than for heterozygous patients (GA) (95%CI: 9% to 124%).

Table 16: Association analysis between single NR1I3 and NR1I2 variants and EFV levels using linear regression models

Gene	rs number	Genomic identification	P-Value	Quantitative effect on EFV levels
NR1I3	rs2307420	10445A>G	0.796	n/a
NR1I3	rs2307424	10396C>T	0.154	n/a
NR1I3	rs34161743	9923C>T	0.708	n/a
NR1I3	rs35205211	9892C>G	0.549	n/a
NR1I3	rs2502815	9774C>T	0.976	n/a
NR1I3	rs3003596	8784T>C	0.579	n/a
NR1I3	rs35709078	7145C>T	0.842	n/a
NR1I3	rs55802895	258G>A	0.531	n/a
NR1I2	rs1523130	5177T>C	0.336	n/a
NR1I2	rs3814055	5705C>T	0.123	n/a
NR1I2	rs1523128	6334G>A	0.017	56% ↑ from 6334AG to 6334AA
NR1I2	rs1523127	6709C>A	0.278	n/a
NR1I2	rs3842689	7070_7075del GAGAAG	0.385	n/a
NR1I2	rs2472677	24087C>T	0.446	n/a
NR1I2	rs12721613	31846C>T	0.160	n/a
NR1I2	rs1464603	32019G>A	0.868	n/a
NR1I2	rs1464602	32042G>A	0.054	n/a
NR1I2	rs2276707	39823C>T	0.956	n/a

↑, increase in EFV levels. ↓, decrease in EFV levels.

4.7.2. The effect of composite variant genotypes on EFV levels

The impact of composite genotypes on EFV levels was examined. Using R statistical software (The R Development Core Team, 2010), linear regression models were used to search for statistically significant composite genotypes that had an effect on EFV levels. Eight *NR1I3* and ten *NR1I2* variants were analysed in conjunction with *CYP2B6*6* (Appendix L). Table 17 shows the mean EFV levels for the *6*6 and Non-*6*6 groups (as explained in Section 3.10.2) in the Xhosa and CMA cohorts. The *NR1I2* and *NR1I3* SNPs were also analysed in conjunction with six SNPs from *CYP1A2*, *CYP2A6*, *CYP3A4* and *CYP3A5* (Appendix L). A total of 12 composite variant pairs were significantly associated with EFV levels ($P \leq 0.05$), which are listed in Table 18. Additive allelic models were used to determine the effect size of these variant pairs on EFV levels.

For *CYP2B6*6*, one composite *NR1I3* SNP, rs55802895 (258G>A) was found to be significantly associated with EFV levels. The 258GG genotype causes an increase in 161% in individuals with two *CYP2B6*6* alleles as opposed to individuals with one or no *CYP2B6*6*-alleles.

Four variants in combination with *CYP1A2*1F* (163C>A) was found to be significantly associated with EFV levels, namely rs2502815, rs3842689, rs12721613 and rs2276707. SNP rs2502815 was found to be associated with increased EFV levels in individuals with the 163CC genotype, however, the sample cohort was very small (n < 10 individuals). The 6-bp deletion (rs3842689) showed a compound effect where individuals heterozygous for *CYP1A2*1F* and negative for the deletion (+/+) (n = 17) showed an increase in EFV levels compared to individuals heterozygous for both *CYP1A2*1F* and the deletion (n = 41). In *NR1I2* rs12721613 (31846C>T), a decrease in EFV levels (86%) is found in individuals with the 31846CT genotype and the 163CC genotype (n = 6) as opposed to the 163AA genotype (n = 11). SNP rs2276707 was also associated with EFV levels in conjunction with *CYP1A2*1F*, however, after the effect size analysis, no discernible pattern of change in EFV levels could be found.

Three composite variants of *CYP2A6*9B* (48T>G) were found to be associated with EFV levels, rs2307420, rs1464603 and rs1464602. The heterozygous genotype for all three variants causes an increase in EFV levels in individuals heterozygous for *CYP2A6*9B*. However, the sample size for these groups was less than 10 individuals.

NR1I2 SNP rs1464602 (32042G>A) was found to be significantly associated with a change in EFV levels in conjunction with *CYP2A6*17* (5065G>A). However, no clear pattern was found when analysing this composite genotype for effect size. In *CYP3A4*1B* (392A>G), *NR1I2* rs1523130 (5177T>C) was found to be associated with decreased EFV levels.

For *CYP3A5*3 (6986A>G)*, two composite variants, *NR1I2* rs3842689 (7070_7075del) and *NR1I2* rs2276707 (39823C>T) were found to be associated with a change in EFV levels. The minor homozygous genotype for both variants were found to cause an increase in EFV levels for individuals heterozygous for *CYP3A5*3*.

Table 17: Mean EFV levels according to CYP2B6*6 genotype for each variant for the Xhosa and CMA cohort

Gene	rs number	Genomic identification	Genotype	Xhosa		CMA	
				Non-*6*6	*6*6	Non-*6*6	*6*6
NR1I3	rs2307420	10445A>G	AA	5.26	13.52	6.80	8.49
			AG	7.64	15.00	5.37	n/a
			GG	n/a	n/a	5.07	n/a
NR1I3	rs2307424	10396C>T	CC	5.65	10.71	5.71	9,32
			CT	4.5	23.45	6.91	n/a
			TT	n/a	n/a	4.72	n/a
NR1I3	rs34161743	9923C>T	CC	5.45	13.89	5.93	8.49
			CT	6.36	n/a	10.00	n/a
			TT	n/a	n/a	n/a	n/a
NR1I3	rs35205211	9892C>G	CC	5.39	13.89	6.13	8.49
			CG	8.15	n/a	1.15	n/a
			GG	n/a	n/a	n/a	n/a
NR1I3	rs2502815	9774C>T	CC	5.79	12.75	5.81	10.46
			CT	4.94	14.58	5.62	6.00
			TT	5.57	n/a	5.24	7.05
NR1I3	rs35709078	7145C>T	CC	5.50	14.96	6.03	8.49
			CT	5.21	6.39	3.98	n/a
			TT	n/a	n/a	n/a	n/a
NR1I3	rs55802895	258G>A	GG	5.43	14.96	5.96	10.68
			GA	5.87	6.39	6.13	1.91
			AA	5.37	n/a	n/a	n/a
NR1I2	rs1523130	5177T>C	TT	5.34	11.45	6.29	12.50
			TC	6.65	16.80	5.48	7.05
			CC	2.12	27.00	5.01	1.91
NR1I2	rs3814055	5705C>T	CC	5.74	16.15	12.5	1,91
			CT	5.64	11.64	7.05	10.68
			TT	3.56	n/a	1.91	n/a
NR1I2	rs1523128	6334A>G	GG	4.77	16.73	6.08	8.49
			GA	6.82	12.19	5.95	n/a
			AA	3.53	n/a	4.93	n/a
NR1I2	rs1523127	6709C>A	CC	5.23	12.71	5.83	12.5
			CA	6.38	11.60	6.40	7.05
			AA	2.87	27.00	5.35	1.91
NR1I2	rs3842689	7070_7075del GAGAAG	+/+	5.63	16.15	5.47	1.91
			+/-	5.52	11.64	6.97	10.68
			-/-	3.89	n/a	12.76	n/a
NR1I2	rs2472677	24087C>T	CC	4.61	14.34	6.97	6.00
			CT	6.31	14.36	5.81	7.05
			TT	5.09	10.70	3.26	10.46
NR1I2	rs12721613	31846C>T	CC	4.96	16.41	6.02	8.49
			CT	7.97	9.70	5.61	n/a
			TT	n/a	n/a	n/a	n/a
NR1I2	rs1464603	32019G>A	GG	5.23	12.41	6.42	12.50
			GA	6.14	16.37	5.17	4.48
			AA	4.08	n/a	n/a	n/a
NR1I2	rs1464602	32042G>A	GG	4.70	14.08	5.75	3.96
			GA	5.77	14.46	6.74	13.03
			AA	7.82	10.70	5.55	n/a
NR1I2	rs2276707	39823C>T	CC	5.79	18.55	6.02	6.53
			CT	5.42	13.53	5.26	1.91
			TT	5.13	6.39	7.67	19.00

Table 18: The effect size (empirical influence on EFV levels) of single and composite NR1I3 and NR1I2 variants that were significantly associated with EFV levels.

Gene	rs number	Genomic identifier	*P*-Value	Effect of genotype on EFV levels
Single variants				
NR1I2	rs1523128	6334A>G	0.017	56% ↑ from 6334AG to 6334AA
Composite variants				
*CYP2B6**6 (516G>T)				
NR1I3	rs55802895	258G>A	0.005	161% ↑ from *6*6|258GG to Non-*6*6|258GG No significant difference from *6*6|258GA to Non-*6*6|258GA
CYP1A2*1F (163C>A)				
NR1I3	rs2502815	9774C>T	0.026	436% ↑ from 163CC|9774CT to 163CC|977CC 390% ↑ from 163CC|9774TT to 163CC|977CC
NR1I2	rs3842689	7070_7075del	0.017	36% ↑ from 163AC|+/- to 163AC|+/+ 263% ↑ from 163AC|-/- to 163AC|+/+
NR1I2	rs12721613	31846C>T	0.026	86% ↓ from 163CC|31846CT to 163AA|31846CT
NR1I2	rs2276707	39823C>T	0.030	No definable pattern detected
CYP2A6*9B (48T>G)				
NR1I3	rs2307420	10445A>G	0.007	464% ↑ from 48 TG|10445AA to 48 TG|10445AG
NR1I2	rs1464603	32019G>A	0.021	215% ↑ from 48TG|32019GG to 48TG|32019AG
NR1I2	rs1464602	32042G>A	0.046	465% ↑ from 48TG|32042GG to 48TG|32042GA
CYP2A6*17 (5065G>A)				
NR1I2	rs1464602	32042G>A	0.003	No definable pattern detected
CYP3A4*1B (392A>G)				
NR1I2	rs1523130	5177T>C	0.011	64% ↑ from 392GG|5177TT to 392GG|5177TC 66% ↓ from 392AG|5177TT to 392AG|5177CC
CYP3A5*3 (6986A>G)				
NR1I2	rs3842689	7070_7075del	0.008	105% ↑ from 6986AG|+/+ to 6986AG|+/- 199% ↑ from 6986AG|+/+ to 6986AG|-/-
NR1I2	rs2276707	39823C>T	0.023	126% ↑ from 6986AG|39823CC to 6986AG|39823TT

↑, increase in EFV levels. ↓, decrease in EFV levels.

4.7.3. The effect of haplotypes on EFV levels

The two variants that were found to be in LD, SNP rs3814055 (5705C>T) and deletion rs3842689, were analysed for an additive haplotype effect on EFV levels. No association was found between any haplotypes for these two variants and EFV levels ($P = 0.463$).

4.8. Correlations between genotype and treatment outcome or adverse reactions

The effect of *NR1I3* and *NR1I2* variants on treatment response was investigated by comparing non-responders to responders in terms of MAFs of each variant. A total of five individuals in the Xhosa cohort and seven individuals in the CMA cohort correspond to criteria for "non-responders" (VL of > 80 copies/ml on date of sample collection). Appendix K shows the MAF of

each variant in the Responder and Non-Responder groups. In the Xhosa cohort, rs3814055 (5705C>T) genotype differed significantly between Responders and Non-Responders (P = 0.011) and in the CMA cohort, rs2472677 (24087C>T) genotype differed significantly between Responders and Non-Responders (P = 0.023).

Figure 14 shows the genotypes for these two SNPs in the respective groups. For both SNPs, more than 50% of individuals in the responder group are homozygous for the major allele (CC), and no individuals in the non-responder group are homozygous for this allele.

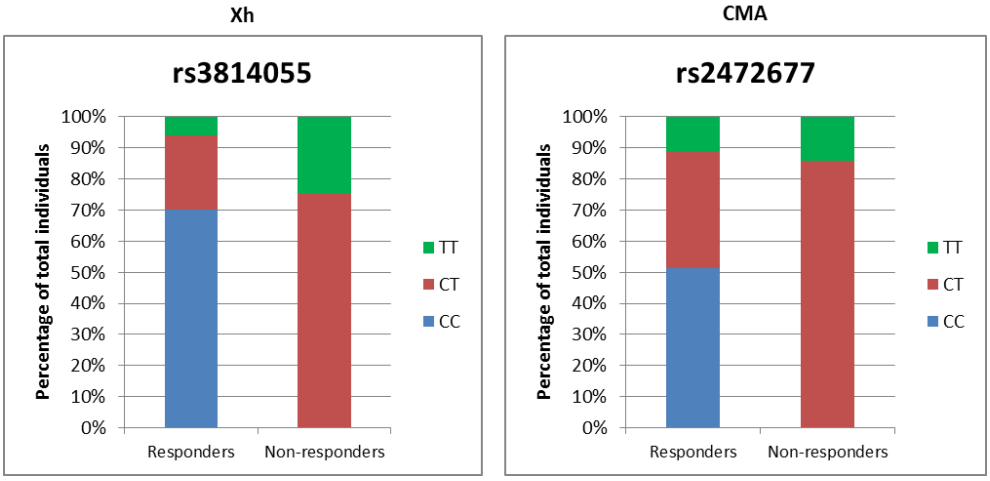


Figure 14: Responder and non-responder genotype frequencies for rs3814055 in the Xhosa cohort and rs2472677 in the CMA cohort

The effect of *NR1I3* and *NR1I2* genotype on the advent of EFV-based ADRs was also analysed. EFV-based ADRs were reported for 11 patients in the Xhosa cohort and five patients in the CMA cohort. The genotype frequencies and MAFs for each variant in the ADR and Non-ADR groups are also shown in Appendix K. No variant differed significantly between the two groups.

Logistic regression analysis of the effect of single variant genotype on treatment outcome and the occurrence of ADRs did not find any statistically significant associations. (Appendix M).

CHAPTER 5

5. Discussion

This study aimed to analyse the effect of *NR1I3* and *NR1I2* variant genotypes on long-term efavirenz levels, treatment outcome and the occurrence of adverse drug reactions. This was achieved through the genotyping of DNA samples collected from HIV-positive female patients of Xhosa and Mixed ancestry who were prescribed an antiretroviral regimen containing efavirenz during the timeframe of the study, and, along with the measured efavirenz levels and available clinical data of the patients, searching for statistically valid correlations.

5.1. Characteristics of study participants and confounding effects

Recruitment of patients for the current study was limited due to the stringent criteria used for the study, including that suitable candidates should be female, of Xhosa or CMA ethnicity, and prescribed an ARV regimen containing EFV for at least three months.

Only female patients were included in the current study. Some studies found an association between gender and EFV levels in the body (Haas *et al*., 2004; Burger *et al*., 2005; Mukonzo *et al*., 2009; Habtewold *et al*., 2011). In these studies female individuals were found to have significantly higher mean EFV plasma levels than male individuals (Burger *et al*., 2005; Mukonzo *et al*., 2009; Habtewold *et al*., 2011). Other studies did not find a significant association, despite specifically testing for the potential confounding effect of gender in the analysis of variants (Marzolini *et al*., 2001; Haas 2004). A possible explanation for the discrepant findings could be that different methods were used to measure EFV levels while time and frequency of measurements also differed between studies. In the current study, the decision to select only female study participants was due to three reasons. Firstly, the current study collaborated with the Women's Interagency HIV Study (WIHS) in the USA and therefore selected females study participants in order to comply with their criteria. Secondly, in order to eliminate variables that could potentially act as confounding effects on EFV levels in the body, only female individuals were selected for our study. Another reason for the exclusion of male individuals was that in the Xhosa population, females were more willing to donate hair samples than male individuals, due to cultural reasons.

Patients participating in the current study were of either Xhosa or CMA ethnicity. One of the aims of this study was to focus specifically on populations relevant to South Africa and the

Western Cape, as HIV is highly prevalent in the country, therefore pharmacogenetic research on ARVs can directly impact this burden in the country. Both the CMA and Xhosa populations are relevant to South Africa, as they are indigenous to the country and constitute a significant portion of the country, especially in the Western Cape (Statistics South Africa, 2013). These two populations also harbour a unique genetic make-up which represents other African population groups as well as contains some European genetic influences, which can provide insight on the differences and similarities among the population groups.

The median age of the Xhosa and CMA participants of the study was 39 and 41 respectively (IQR: 34 – 45 and 37 – 46 respectively). The effect of patient age had variable results in studies searching for confounders influencing EFV levels. Amoroso *et al*. (2012) found a significant association between older age and increased odds of viral suppression (decreased VL). Prosperi *et al*. (2012) found a significant association with older age and an increased risk of EFV-based ADRs. In other studies, however, no association was found between age and any EFV-linked outcome (Marzolini *et al*., 2001; Motsinger *et al*., 2006; Ritchie *et al*., 2006; Gandhi *et al*., 2009; Wyen *et al*., 2011; Sukasem *et al*., 2013). A possible explanation for the discrepancy between different studies could be the difference in the intervals between VL measurements and the total number of measurements taken. In the current study, age was significantly associated with treatment outcome as indicated by VL (P = 0.0270) (Figure 15), correlating with the Amoroso *et al*., (2012) study.

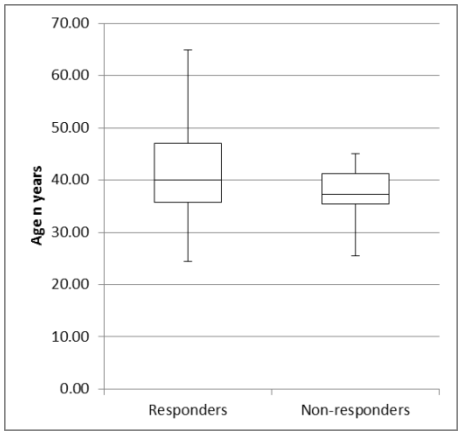


Figure 15: Box plot for age for Responders (VL < 80 copies/ml) and Non-responders (V > 80 copies/ml).

BMI and weight has previously been associated with EFV-related outcomes in HIV-positive patients. Evans *et al*. (2012) reported that a low BMI was significantly associated with negative

treatment outcome of EFV or NVP as indicated by a VL > 400 copies/ml. Other studies also found an association between weight and EFV plasma levels (Arab-Alameddine *et al*., 2009; Di Iulio *et al*., 2009). It was found that an increased weight was associated with higher EFV clearance, and, therefore, decreased EFV plasma levels (Arab-Alameddine *et al*., 2009; Di Iulio *et al*., 2009). The current study did not find an association with weight and EFV levels, or any other outcome. It is possible that BMI and weight effect are ethnic-specific. The individuals that participated in the three studies mentioned above were predominantly Caucasian, whereas the current study did not include any individuals of Caucasian ethnicity. It should also be mentioned that by using BMI instead of weight as a variable for this study, the effect of height on body mass would have been taken into account, and could have resulted in a different outcome.

In the current study, time using EFV was found to be significantly associated with the occurrence of reported EFV-based ADRs. Individuals that reported one or more occurrences of EFV-based ADRs were prescribed EFV for a longer period than individuals with no reported occurrences of EFV-based ADRs (Figure 16). This could be due to the fact that the longer time an individual was prescribed EFV, the longer time is available to report the occurrence of ADRs. The reporting of ADRs is largely done verbally by the patients' themselves, which creates the possibility that not all ADRs are reported and some instances may be overlooked by the clinician.

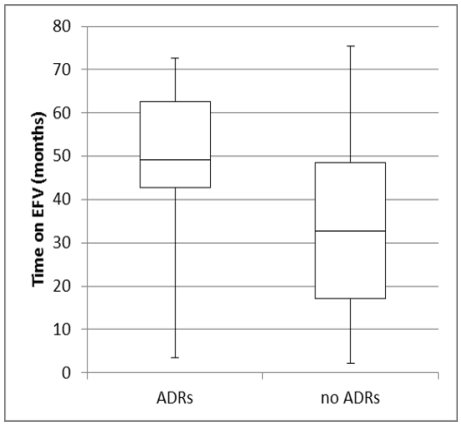


Figure 16: Box plot for time using EFV for individuals with reported EFV-based ADRs and individuals with no reported EFV-based ADRs.

5.2. Survey of sequence conservation in *NR1I3*

In silico analysis was performed in order to find putative conserved regions in the *NR1I3* gene and its orthologs in various species. From mVISTA results, three conserved regions were highlighted: a 260 bp region just upstream of exon 1, the exon 1-region and the exon 2-region

(Figure 10). Due to the occurrence of enzyme slippage, bi-directional sequencing results for the 260 bp upstream region was unsuccessful.

Sequencing results indicate that the exon 1- and exon 2-regions, also including the first 40 bp and last 56 bp of intron 1, were conserved within the 30-patient sequencing cohort, as bi-directional sequencing results identified no variants in these regions. This was expected, as the occurrence of base pair variants in this region can result in the alteration of transcription and translation of *NR1I3*, and therefore the genetic conservation of these areas are important. The 5'-UTR region of *NR1I3*, of which exon 1 forms part, plays a role in the transcriptional regulation of *NR1I3*, and a variant occurring in this region can result in the alteration of mRNA secondary structure, TFBS and ribosomal binding sites, potentially influencing the transcription rate of *NR1I3*. An exonic variant can result in an amino acid change when the gene is translated, potentially altering protein activity. A variant occurring in the 50-bp intronic regions flanking the exons, could potentially disrupt an exonic splice site.

The promoter area of *NR1I3* was also analysed for the presence of putative TFBS as identified by prediction software, as well as functional TFBS reported by other studies. It was found that only one SNP from sequencing results (rs55802895, 258G>A) disrupted a putative TFBS as identified in AliBaba (RAR-α), and no SNPs disrupted any of the functionally identified TFBS (Table 10). This shows that the TFBS in conserved, emphasising the importance of transcriptional regulation of *NR1I3*, even though it is constitutively expressed.

From bi-directional sequencing analysis, it was found that the *NR1I3* exonic regions were well conserved among the 30 individuals sequenced. Only 12 SNPs were found in the sequenced amplicons, which consisted of 738 bp 5'-upstream region, 393 bp UTR (including exon 1), 1,080 bp exonic region and 1,754 intronic region (Table 11). Only three SNPs out of the 12 occurred in the exonic regions, two SNPs in exon 4 and one SNP in exon 5. None of the SNPs caused an amino acid change.

In most nuclear receptors, the DBD should be more conserved than the LBD, both in terms of protein and DNA contents, in order for CAR to recognise and bind to response elements in target genes (Maglich *et al*., 2001; Wang & LeCluyse, 2003; Lamba 2008). In terms of number of SNPs found in the sequencing results, the DBD was more conserved than the LBD of *NR1I3* in the current study. The DBD is 682 bp in length, and was included entirely in the sequenced amplicons. The total length of the LBD is 3,458 bp, but only 1,646 bp was included in the sequenced amplicons. From sequencing results, only one SNP, rs35709078, was identified in the DBD region, which resulted in a SNP density for the DBD of 1 SNP per 682 bp. Five SNPs were

identified in the LBD, which resulted in a SNP density of one SNP per 329 bp, confirming that the LBD is less conserved than the DBD. However, the results obtained would be viewed with more confidence if the number of individuals sequenced is increased, and if the entire LBD region is sequenced.

5.3. Comparison of variant genotype and allele frequency in the study cohort and large database populations

From the *in silico* analysis, bi-directional sequencing results, and literature analysis, target *NR1I3* SNPs were selected and subsequently genotyped. For *NR1I2*, analysis of literature, as well as the 1000 Genomes and HapMap databases, resulted in the identification of *NR1I2* target variants for subsequent genotyping.

One SNP, rs1523130 (5177T>C) was genotyped using a Taqman® Genotyping Assay. Two samples, EF141 and EF152, failed on the Taqman® genotyping assay. Sample EF141 showed no fluorescence at all and was therefore not genotyped. The lack of fluorescence could be explained by the presence of a possible novel mutation within the recognition site of the Taqman® probe, which would result in the inability for the probe to bind, or could be an indication of the presence of inhibitors of the PCR reaction. Sample EF152 showed some fluorescence for both probes, but the fluorescence was too low for genotyping. The low level of fluorescence could be explained by the presence of inhibitors of the PCR reaction, or a low sample DNA concentration.

From genotyping results for both genes, it was found that *NR1I3* SNPs generally occurred at a lower frequency than *NR1I2* variants. Seven out of the nine *NR1I3* SNPs in the Xhosa cohort and six out of nine *NR1I3* SNPs in the CMA cohort had a MAF lower than 0.100 (Table 13). This concurs with other studies that also reported low MAFs for most *NR1I3* SNPs (Ikeda *et al*., 2003; Thompson *et al*., 2005; Chew *et al*., 2013). Seven out of nine *NR1I2* variants in both cohorts had a MAF of higher than 0.200 (Table 13). This is in accordance with the observation that *NR1I3* is more genetically conserved than *NR1I2*. The high conservancy of *NR1I3* as compared to *NR1I2* can be explained by the larger size of *NR1I2*, creating more opportunity for genetic variants to occur.

The average MAF for intronic variants was 0.197 (n = 4) in *NR1I3* and 0.340 (n = 4) in *NR1I2*, and for exonic SNPs was 0.038 (n = 3) in *NR1I3* and 0.067 (n = 1) in *NR1I2*. For both *NR1I3* and *NR1I2* variants, it was observed that the intronic variants had a higher MAF than the exonic SNPs, which concurs with the observation that exons are genetically more conserved than introns.

When comparing the Xhosa and CMA cohorts in terms of frequency distribution of SNPs, it was found that the MAFs of one *NR1I3* SNP (rs2307424) and three *NR1I2* SNPs (rs1523130, rs3814055 and rs1523127) differed significantly between the two cohorts ($P \leq 0.05$). For three of these SNPs (rs2307424, rs1523130 and rs1523127), the CMA cohort had a higher MAF than the Xhosa cohort, and for one SNP (rs3814055), the Xhosa MAF was higher. The occurrence of such a low number of variants of significantly different MAFs, could indicate that the Xhosa and CMA cohorts are closely related. This correlated with a study by Ikediobi *et al*. (2011), who analysed approximately 200 SNPs occurring in pharmacogenetically relevant genes in the Xhosa and CMA populations, and found eight SNPs that differed significantly with regards to MAF. This is expected, as the Xhosa and CMA populations share N-K ancestry (Tishkoff *et al*., 2009).

Subsequently, the MAFs for all target SNPs in the two cohorts were compared to MAFs of five populations from the 1000 Genomes database and one population from the HapMap database. The results indicate that the CEU, CHB and GIH populations had the highest difference in MAFs from the Xhosa and CMA cohorts (Table 14). This is expected, as the two South African cohorts would share similar genetic information with geographically closer cohorts, like the African cohorts YRI and LWK. The ASW population is also expected to be similar in genetic structure to both the Xhosa and CMA cohorts, as this population shares N-K ancestry with both the Xhosa and CMA cohorts, as well as European Caucasian ancestry with the CMA cohort (Tishkoff *et al*., 2009). The SNP MAFs for CEU and GIH populations were closer to the CMA frequency distribution than the Xhosa population, as they shared a higher percentage similar SNP MAFs (Table 14). This was also expected, as the CMA population contains some European Caucasian and Indian ancestry (Tishkoff *et al*., 2009). The Xhosa and CMA cohorts showed equal differences in MAFs to the CHB population, which is expected, as neither cohort contains a high amount of East Asian ancestry (Tishkoff *et al*., 2009). A high percentage of SNPs differed in MAF between the YRI population and the study cohorts. This is unexpected for the Xhosa cohort, as the genetic constituency of the Xhosa population shares some similarities with populations from Central and West Africa, of which the Yoruban population forms part (Tischkoff *et al*. 2009; Schuster *et al*. 2010). A possible explanation for this occurrence could be that the Xhosa study cohort might not be fully homogenous and could contain some admixture (Patterson *et al*., 2010).

All variants except *NR1I3* SNPs rs113800307 and rs35709078 were submitted to LD analysis using Haploview v4.2 (Barrett *et al*., 2005). No LD was found among the *NR1I3* SNPs. This was unexpected, because, in a previous study (Oliver *et al*., 2010), *NR1I3* SNPs rs2307424 and

rs2502815 were found to be in LD. Only two instances of LD were found in *NR1I2*; rs3814055 in the 5'-upstream region was found to be in LD with rs3842689 in the 5'-UTR region and rs1523130 in the 5'-upstream region was found to be in LD with rs1523127 in the 5'-UTR region. In another study (Fanta *et al*., 2008), a haplotype was found consisting of five variants, rs3814055, rs1523127, rs3842689, rs6785049 and rs2276707, associated with cyclosporine availability. However, in the current study, rs6785049 was not genotyped and rs2276707 was not linked to the other variants. It can be suggested that the LD results in the current study differ from previous studies, because of differences in ethnicity of the study participants, as the studies mentioned above used mainly Caucasian participants.

5.4. The effect of variant genotype on EFV levels

The effect of genetic variation in *NR1I3* and *NR1I2* on EFV pharmacokinetics as measured by EFV levels in hair was analysed, first through preliminary analysis using Microsoft Office Excel 2010, followed by linear regression analysis using R statistical software (The R Development Core Team, 2010). The variation in the two nuclear receptor genes were analysed independently, as well as in combination with variants in genes encoding for EFV-metabolising proteins.

From the statistical analysis results, *NR1I2* SNP rs1523128 (6334A>G) was found to be associated with EFV hair levels. Linear regression models predict the effect size of the SNP to be a decrease of 56% in EFV in individuals homozygous for the A-allele (AA) compared to heterozygotes (GA) (Table 17). This SNP was also important in the preliminary analysis, where minor homozygotes (AA) for this SNP had the lowest mean EFV levels in the Xhosa cohort (Table 15). SNP rs1523128 occurs in the 5'-upstream region of *NR1I2*, 893 bp from the start of exon 1. Previous *in silico* analyses found that the rs1523128 A-allele creates a putative C/EBP TFBS and disrupts a putative HNF-1α TFBS (Zhang *et al*., 2001; Lamba *et al*. 2008). The C/EBP TF has been found to cooperate with both CAR and PXR in transcriptional regulation of xenobiotic metabolising genes, including *CYP*-genes (Goodwin *et al*., 2002; Benet *et al*., 2010). HNF-1α is also involved in the transcriptional regulation of CYP-genes as well as the NR farnesoid X receptor (FXR) (Gonzalez & Lee, 1996; Vacca *et al*., 2013). However, the transcriptional regulation of PXR by the two TFs has yet to be investigated. It can be hypothesised that the disruption of a TFBS and the creation of another alters the activation and subsequent rate of transcription of PXR. This could lead to an increase in the induction of *CYP2B6* and other EFV-metabolising proteins by PXR, which in turn could lead to lower EFV levels. SNP rs1523128 was mentioned in two other studies, where it was associated with susceptibility to Crohn's Disease (inflammatory bowel disorder) and periodontitis (inflammatory

oral cavity disorder) (Glas *et al*., 2011; Folwaczny *et al*., 2012). However, this SNP was not associated with EFV levels in any other study.

5.4.1. The effect of composite genotype on EFV levels

NR1I3 SNP rs55802895 (258G>A), in conjunction with *CYP2B6*6*, was significantly associated with EFV levels. Individuals with the *6*6|258GG composite genotype showed a 161% increase in EFV levels compared to individuals with the Non-*6*6|258GG composite genotype (Table 17). Individuals with the 258GA or 258AA genotype showed no significant change in mean EFV levels between the **6*6* and Non-**6*6* groups (Table 17). SNP rs55802895 occurs in the 5'-upstream region of *NR1I3*, and the A-allele disrupts a putative RAR-α TFBS, as predicted using the AliBaba *in silico* program (Grabe, 2002). Saito *et al.* (2010) found that the binding of RAR TF to a TFBS in intron 1 of *NR1I3* increases the transcription of the gene. It can therefore be postulated that the disruption of the putative TFBS by the rs55802895 minor allele could result in decreased *NR1I3* transcription, resulting in decreased *CYP2B6* activation and subsequent increased EFV levels. However, the opposite is observed, namely that the presence of the rs5580295 minor allele (A-allele) reduces the effect of *CYP2B6*6* on EFV levels, resulting in lower EFV levels.

The genotyped *NR1I3* and *NR1I2* genes have also been analysed for their composite effect on variants in four other genes involved in EFV metabolism, *CYP1A2*, *CYP2A6*, *CYP3A4* and *CYP3A5*. A range of significant associations of *NR1I3* and *NR1I2* variants were identified, that exerted either a reducing or an amplifying effect on the influence of their composite SNP on EFV levels (Table 17).

SNP *CYP1A2*1F* induces increased activity of the CYP1A2 enzyme (Han *et al*., 2002), which could result in decreased drug levels. This SNP formed a compound genotype with one *NR1I3* SNP (rs2502815) and three *NR1I2* SNPs (rs3842689, rs12721613 and rs2276707) which influenced EFV hair levels in the study cohort. *NR1I3* SNP rs2502815 showed a trend in decreased EFV levels in a study by Swart *et al*. (2012), although not statistically significant. *NR1I2* SNP was associated with decreased PXR activity (Zhang *et al*., 2001). *CYP2A6*9B* was associated with elevated EFV plasma levels in a previous study (Kwara *et al*., 2009). Compound genotypes were identified of this SNP along with one *NR1I3* SNP (rs2307420) and two *NR1I2* SNPs (rs1464603 and rs1464602). Individuals with these compound genotypes had increased EFV levels, which indicates that the *NR1I3* and *NR1I2* SNPs had an amplifying effect on *CYP2A6*9B*. *CYP2A6*17* was also associated with higher mean EFV plasma levels in a previous

study (Kwara *et al*., 2009). One *NR1I2* SNP was found to influence EFV levels in conjunction with *CYP2A6*9B*, however, no discernible pattern could be detected. The impact of *CYP3A4*1B* on CYP3A4 expression is unclear, and varying results are produced by different studies (Wojnowski, 2004; Klein & Zanger, 2013), indicating the possible impact of other additional SNPs on *CYP3A4*1B*. In the current study, *NR1I2* SNP rs1523130 was found to influence EFV levels in conjunction with *CYP3A4*1B*. *CYP3A5*3* creates a splice error, which leads to a decrease in CYP3A5 expression. Two *NR1I2* SNPs, rs3842689 and rs2276707, were found to have an influence on EFV levels in conjunction with *CYP3A5*3*. These results emphasise the complexity of EFV metabolism and the various effects of all the elements involved.

5.5. The effect of variant genotype on treatment outcome and adverse drug reactions

The effect of genetic variation in *NR1I3* and *NR1I2* and treatment outcome and the occurrence of ADRs were analysed, first through preliminary analysis, and secondly through logistic regression models using R statistical software (The R Development Core Team, 2010). No variants were associated with treatment outcome or the occurrence of ADRs. The number of non-responders in the study cohort was small (n = 5 and n = 7 for the Xhosa and CMA cohorts respectively). The number of individuals with reported ADRs was also small (n = 11 and n = 5 for the Xhosa and CMA cohorts respectively). This small sample size limits the validity and strength of the statistical analysis. Significant statistical findings may occur in a cohort with a larger portion of non-responders and individuals with reported ADRs.

CHAPTER 6

6. Conclusions and future research

This study aimed to investigate the genetic variants in two nuclear receptor genes in two South African populations, and to analyse the effect of these on EFV levels and treatment outcome. Based on this aim, the following conclusions were drawn:

The analysis of the characteristics of the study participants identified statistically significant confounding effects, which were adjusted for in the analysis:

- Age was significantly associated with treatment outcome, as the mean age for responders was higher than the mean age of non-responders.
- BMI and weight was not significantly associated with any of the study outcomes.
- Time on efavirenz was significantly associated with the occurrence of ADRs, as patients that were prescribed EFV for a longer period of time had a higher occurrence of reported ADRs.

The focus on the two nuclear receptors highlighted the importance of transcriptional regulation, and showed the need for more in-depth investigation into the activation and de-activation of transcriptional regulators, and the downstream effects of this action:

- In *NR1I3*, the DBD was more conserved than the LBD when analysing the sequencing results, which confirmed *in silico* analysis.
- In *NR1I3*, exonic SNPs occurred at a lower MAF than intronic SNPs, which confirms selective constraint.
- *NR1I3* SNPs have a lower MAF in general than *NR1I2* variants, which confirms *in silico* analysis and literature proclaiming that *NR1I3* is more conserved.

Through the investigation of two important South African populations, and the comparison of the genetic variants in these two cohorts with other populations, it was shown that these populations share some common genetic composition with each other, as well as with other African populations, but also showed some significant differences in genetic variation:

- Only four out of the 19 genotyped variants in total differed significantly between the Xhosa and CMA cohorts, which confirm the suggestion that these two populations share some genetic patterns.

- When comparing the variant MAFs with five other populations form HapMap and 1000 Genomes databases, it was found that the Xhosa cohort is most similar to the LWK and ASW population and the CMA cohort is most similar to the ASW population. The CEU, CHB and GIH populations are the least similar to both study cohorts.

Through the analysis of genetic variation in two important nuclear receptor genes, their effect in conjunction with primary EFV-metabolizing gene variants and their influence on EFV levels, treatment outcome and the occurrence of adverse reactions, the complexity of EFV metabolism was emphasised.

- After correcting for confounders, *NR1I3* SNP rs1523128 was associated with decreased EFV levels (P = 0.0168) in both cohorts. In individuals homozygous for the minor allele of rs1523128, a decrease of 56% in EFV levels was observed.
- Individuals with *CYP2B6**6*6 genotype and *NR1I3* SNP rs55802895 GG genotype showed significant increase in EFV levels, which leads to the conclusion that this genotype has the ability to enhance the effect of *CYP2B6**6*6.
- Following analysis of *NR1I2* and *NR1I3* SNPs, in combination with SNPs from *CYP1A2*, *CYP2A6*, *CYP3A4* and *CYP3A5*, 11 composite genotypes were identified that were significantly associated with increased or decreased EFV levels (Table 17).
- No variants showed a statistically significant effect on treatment outcome or the occurrence of adverse drug reactions. This can be explained by the small portion of the cohort that was identified as bad responders or that reported the occurrence of ADRs.
- In the attempt to attribute significant effects on EFV treatment outcome and the occurrence of ADRs to genetic variance in the two nuclear receptor genes, it was concluded that more elaborate and standardised measurements of treatment outcome and reports of ADRs are needed in order to gain valid results.

The findings of this study prompt the further examination of the EFV metabolism pathway, and some factors should be carefully considered:

- In the current study, sample size was a limiting factor. The small sample size of the study lowers the statistical impact of the results, as it increases the odds of significant results occurring due to chance. By dividing the sample cohort into two groups based on ethnicity, further decreases the sample size. It should be a priority of future studies to collect a sample cohort with as high as possible cohort size, which would ensure that statistical results are trustworthy and have a high power. The participation of more clinics

in future studies can help to increase the sample size, however, due to the high risk of stratification in the CMA population, the possibility increases of inadvertently merging CMA subgroups together that are in fact genetically separate entities.

- The fact that the current sample cohort was started on ART at different times, differed in terms of regimen history and the recording of VL-measurements and CD4 counts were not at standardised times, complicated the design of statistical assays. Therefore, future research studies should ensure that clinical history for each patient is uniform, complete, up to date and standardised. This is currently being addressed in a new study in our group.
- In order to be able to analyse the genetic conservation in both nuclear receptor genes equally, it would have been favourable to sequence the exonic region of the *NR1I2* gene as was done for the *NR1I3*, and is therefore recommended for further research.
- It would be helpful to investigate the downstream effect of *NR1I3* and *NR1I2* regulation on a protein level, as an addition to genetic analysis, by looking at the inhibition or activation of protein activity.
- Another aspect of EFV-metabolism is the transport of the ARV drug and its metabolites in and out of cells. The effect of genetic variation in transporters linked to EFV, for example *ABCB1*, should be further investigated. This forms part of another study in our group.

Using, among others, the above suggestions, the mechanism of EFV metabolism could be fully unravelled, which would aid in optimisation and management of patient ARV regimen dosage. This could help reduce the occurrence of adverse effects, as well as increase the chances of treatment success in patients.

References

Adkins, J. C., & Noble, S. (1998). Efavirenz. *Drugs*, *56*(6), 1055–1064.

Almond, L. M., Hoggard, P. G., Edirisinghe, D., Khoo, S. H., & Back, D. J. (2005). Intracellular and plasma pharmacokinetics of efavirenz in HIV-infected individuals. *The Journal of Antimicrobial Chemotherapy*, *56*(4), 738–744.

Alnouti, Y., & Klaassen, C. D. (2008). Regulation of sulfotransferase enzymes by prototypical microsomal enzyme inducers in mice. *The Journal of Pharmacology and Experimental Therapeutics*, *324*(2), 612–621.

Amoroso, A., Etienne-Mesubi, M., Edozien, A., Ojoo, S., Sheneberger, R., Obiefune, M., … Redfield, R. R. (2012). Treatment outcomes of recommended first-line antiretroviral regimens in resource-limited clinics. *Journal of Acquired Immune Deficiency Syndromes*, *60*(3), 314–320.

Amunugama, H. T., Zhang, H., & Hollenberg, P. F. (2012). Mechanism-based inactivation of cytochrome P450 2B6 by methadone through destruction of prosthetic heme. *Drug Metabolism & Disposition*, *40*(9), 1765–1770.

Andrews, E., Armstrong, M., Tugwood, J., Swan, D., Glaves, P., Pirmohamed, M., … Daly, A. K. (2010). A role for the pregnane X receptor in flucloxacillin-induced liver injury. *Hepatology*, *51*(5), 1656–1664.

Antonelli, G., & Turriziani, O. (2012). Antiviral therapy: Old and current issues. *International Journal of Antimicrobial Agents*, *40*(2), 95–102.

Arab-Alameddine, M., Di Iulio, J., Buclin, T., Rotger, M., Lubomirov, R., Cavassini, M., … Csajka, C. (2009). Pharmacogenetics-based population pharmacokinetic analysis of efavirenz in HIV-1-infected individuals. *Clinical Pharmacology and Therapeutics*, *85*(5), 485–494.

Arnold, K. A., Eichelbaum, M., & Burk, O. (2004). Alternative splicing affects the function and tissue-specific expression of the human constitutive androstane receptor. *Nuclear Receptor*, *2*(1), 1.

Asimus, S., & Ashton, M. (2009). Artemisinin: A possible *CYP2B6* probe substrate? *Biopharmaceutics & Drug Disposition*, *30*, 265–275.

Assenat, E., Gerbal-Chaloin, S., Larrey, D., Saric, J., Fabre, J.-M., Maurel, P., … Pascussi, J. M. (2004). Interleukin 1beta inhibits CAR-induced expression of hepatic genes involved in drug and bilirubin clearance. *Hepatology (Baltimore, Md.)*, *40*(4), 951–960.

Avery, L. B., Sacktor, N., McArthur, J. C., & Hendrix, C. W. (2013). Protein-free efavirenz concentrations in cerebrospinal fluid and blood plasma are equivalent: applying the law of mass action to predict protein-free drug concentration. *Antimicrobial Agents and Chemotherapy*, *57*(3), 1409–1414.

Bangsberg, D. R., Acosta, E. P., Gupta, R., Guzman, D., Riley, E. D., Harrigan, P. R., … Deeks, S. G. (2006). Adherence-resistance relationships for protease and non-nucleoside reverse transcriptase inhibitors explained by virological fitness. *AIDS*, *20*(2), 223–231.

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*(2), 263–265.

Barril, P., & Nates, S. (2012). Introduction to agarose and polyacrylamide gel electrophoresis matrices with respect to their detection sensitivities. In S. Magdeldin (Ed.), *Gel Electrophoresis - Principles and Basics* (1st ed., pp. 3–15). Rijeka, Croatia: InTech.

Bates, I., Fenton, C., Gruber, J., Lalloo, D., Lara, A. M., Squire, S. B., … Tolhurst, R. (2004). Vulnerability to malaria, tuberculosis, and HIV / AIDS infection and disease. Part 1: Determinants operating at individual and household level. *Lancet Infectious Diseases*, *4*, 267–277.

Baye, T. M., & Wilke, R. A. (2010). Mapping genes that predict treatment outcome in admixed populations. *The Pharmacogenomics Journal*, *10*(6), 465–477.

Belanger, A.-S., Caron, P., Harvey, M., Zimmerman, P. A., Mehlotra, R. K., & Guillemete, C. (2009). Glucuronidation of the antiretroviral drug efavirenz by UGT2B7 and an *in vitro* investigation of drug-drug interaction with zidovudine. *Drug Metabolism and Disposition*, *37*(9), 1793–1796.

Bera, E., & Mia, R. (2012). Safety of nevirapine in HIV-infected pregnant women initiating antiretroviral therapy at higher CD4 counts: A systematic review and meta-analysis. *South African Medical Journal*, *102*(11), 855–559.

Bertilsson, G., Heidrich, J., Svensson, K., Asman, M., Jendeberg, L., Sydow-Bäckman, M., … Berkenstam, A. (1998). Identification of a human nuclear receptor defines a new signaling pathway for *CYP3A* induction. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(21), 12208–12213.

Bertrand, S., Brunet, F. G., Escriva, H., Parmentier, G., Laudet, V., & Robinson-Rechavi, M. (2004). Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. *Molecular Biology and Evolution*, *21*(10), 1923–1937.

Best, B. M., Koopmans, P. P., Letendre, S. L., Capparelli, E. V, Rossi, S. S., Clifford, D. B., … Grant, I. (2011). Efavirenz concentrations in CSF exceed IC50 for wild-type HIV. *The Journal of Antimicrobial Chemotherapy*, *66*(2), 354–357.

Boulle, A. (2008). Antiretroviral therapy and early mortality in South Africa. *Bulletin of the World Health Organization*, *86*(9), 678–687.

Bristol-Myers Squibb Pharma Company. (2013). Sustiva® (Efavirenz) Prescribing Information. Retrieved September 12, 2013, from http://packageinserts.bms.com/pi/pi_sustiva.pdf

Brown, K. C., Hosseinipour, M. C., Hoskins, J. M., Thirumaran, R. K., Tien, H., Weigel, R., … Corbett, A. H. (2012). Exploration of *CYP450* and drug transporter genotypes and correlations with nevirapine exposure in Malawians. *Pharmacogenomics*, *13*(1), 113–121.

Burger, D., Van der Heiden, I., La Porte, C., Van der Ende, M., Groeneveld, P., Richter, C., … Van Schaik, R. (2005). Interpatient variability in the pharmacokinetics of the HIV non-nucleoside reverse transcriptase inhibitor efavirenz: the effect of gender, race, and *CYP2B6* polymorphism. *British Journal of Clinical Pharmacology*, *61*(2), 148–154.

Burhenne, J., Matthée, A.-K., Pasáková, I., Röder, C., Heinrich, T., Haefeli, W. E., … Weiss, J. (2010). No evidence for induction of ABC transporters in peripheral blood mononuclear cells in humans after 14 days of efavirenz treatment. *Antimicrobial Agents and Chemotherapy*, *54*(10), 4185–4191.

Burk, O., Koch, I., Raucy, J., Hustert, E., Eichelbaum, M., Brockmöller, J., … Wojnowski, L. (2004). The induction of cytochrome P450 3A5 (CYP3A5) in the human liver and intestine is mediated by the xenobiotic sensors pregnane X receptor (PXR) and constitutively activated receptor (CAR). *The Journal of Biological Chemistry*, *279*(37), 38379–38385.

Burk, O., Arnold, K. A., Geick, A., Tegude, H., & Eichelbaum, M. (2005). A role for constitutive androstane receptor in the regulation of human intestinal *MDR1* expression. *Biological Chemistry*, *386*, 503–513.

Castelnuovo, B., Sempa, J., Agnes, K. N., Kamya, M. R., & Manabe, Y. C. (2011). Evaluation of WHO criteria for viral failure in patients on antiretroviral treatment in resource-limited settings. *AIDS Research and Treatment*, 736938.

Castillo, R., Pedalino, R. P., El-Sherif, N., & Turitto, G. (2002). Efavirenz-associated QT prolongation and Torsade de Pointes arrhythmia. *The Annals of Pharmacotherapy*, *36*(6), 1006–1008.

Chang, T. K. H., Bandiera, S. M., & Chen, J. (2003). Constitutive androstane receptor and pregnane X receptor gene expression in human liver: Interindividual variability and correlation with *CYP2B6* mRNA levels. *Drug Metabolism and Disposition*, *31*(01), 7–10.

Chekmenev, D. S., Haid, C., & Kel, A.E. (2005). P-Match: Transcription factor binding site search by combining patterns and weight matrices. *Nucleic Acids Research*, *33*, W432–W437.

Chen, Y., Kissling, G., Negishi, M., & Goldstein, J. A. (2005). The nuclear receptors constitutive androstane receptor and pregnane X receptor cross-talk with hepatic nuclear factor 4α to synergistically activate the human *CYP2C9* promoter. *The Journal of Pharmacology and Experimental Therapeutics*, *314*(3), 1125–1133.

Chen, Y., Tang, Y., Guo, C., Wang, J., Boral, D., & Nie, D. (2012). Nuclear receptors in the multidrug resistance through the regulation of drug-metabolizing enzymes and drug transporters. *Biochemical Pharmacology*, *83*(8), 1112–1126.

Cortes, C. P., Siccardi, M., Chaikan, A., Owen, A., Zhang, G., & La Porte, C. J. L. (2013). Correlates of efavirenz exposure in Chilean patients affected with human immunodeficiency virus reveals a novel association with a polymorphism in the constitutive androstane receptor. *Therapeutic Drug Monitoring*, *35*(1), 78–83.

De Wit, E., Delport, W., Rugamika, C. E., Meintjes, A., Möller, M., van Helden, P. D., … Hoal, E. G. (2010). Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Human Genetics*, *128*(2), 145–153.

Di Iulio, J., Fayet, A., Arab-Alameddine, M., Rotger, M., Lubomirov, R., Cavassini, M., … Telenti, A. (2009). In vivo analysis of efavirenz metabolism in individuals with impaired CYP2A6 function. *Pharmacogenetics and Genomics*, *19*(4), 300–309.

Di Masi, A., De Marinis, E., Ascenzi, P., & Marino, M. (2009). Nuclear receptors CAR and PXR: Molecular, functional, and biomedical aspects. *Molecular Aspects of Medicine*, *30*(5), 297–343.

Durant, J., Clevenbergh, P., Garraffo, R., Halfon, P., Icard, S., Del Giudice, P., … Dellamonica, P. (2000). Importance of protease inhibitor plasma levels in HIV- infected patients treated with genotypic-guided therapy: pharmacological data from the Viradapt Study. *AIDS*, *14*(10), 1333–1339.

Duret, C., Daujat-Chavanieu, M., Pascussi, J.-M., Pichard-Garcia, L., Balaguer, P., Fabre, J.-M., … Gerbal-Chaloin, S. (2006). Ketoconazole and miconazole are antagonists of the human glucocorticoid receptor: Consequences on the expression and function of the constitutive androstane receptor and the pregnane X receptor. *Molecular Pharmacology*, *70*(1), 329–339.

Duval, X., Peytavin, G., Breton, G., Ecobichon, J.-L., Descamps, D., Thabut, G., & Leport, C. (2007). Hair versus plasma concentrations as indicator of indinavir exposure in HIV-1-infected patients treated with indinavir/ritonavir combination. *AIDS*, *21*(1), 106–108.

Ekouevi, D. K., Coffie, P. A., Ouattara, E., Moh, R., Amani-Bosse, C., Messou, E., … Dabis, F. (2011). Pregnancy outcomes in women exposed to efavirenz and nevirapine: An appraisal of the IeDEA West Africa and ANRS databases, Abidjan, Côte d'Ivoire. *Journal of Acquired Immune Deficiency Syndromes*, *56*(2), 183–187.

Evans, D., Maskew, M., & Sanne, I. (2012). Increased risk of mortality and loss to follow-up among HIV-positive patients with oropharyngeal candidiasis and malnutrition before antiretroviral therapy initiation: A retrospective analysis from a large urban cohort in Johannesburg, South Africa. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, *113*(3), 362–372.

Fairall, L., Bachmann, M. O., Lombard, C., Timmerman, V., Uebel, K., Zwarenstein, M., … Bateman, E. (2012). Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH): A pragmatic, parallel, cluster-randomised trial. *Lancet*, *380*(9845), 889–898.

Fanta, S., Jönsson, S., Karlsson, M. O., Niemi, M., Holmberg, C., Hoppu, K., & Backman, J. T. (2010). Long-term changes in cyclosporine pharmacokinetics after renal transplantation in children: Evidence for saturable presystemic metabolism and effect of *NR1I2* polymorphism. *Journal of Clinical Pharmacology*, *50*(5), 581–597.

Faucette, S. R., Zhang, T., Moore, R., Sueyoshi, T., Omiecinski, C. J., Lecluyse, E. L., … Wang, H. (2007). Relative activation of human pregnane X receptor versus constitutive androstane receptor defines distinct classes of *CYP2B6* and *CYP3A4* inducers. *The Journal of Pharmacology and Experimental Therapeutics*, *320*(1), 72–80.

Fellay, J., Marzolini, C., Meaden, E. R., Back, D. J., Buclin, T., Chave, J. P., … Telenti, A. (2002). Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: A pharmacogenetics study. *Lancet*, *359*(9300), 30–36.

Ferradini, L., Jeannin, A., Pinoges, L., Izopet, J., Odhiambo, D., Mankhambo, L., … Brasher, C. (2006). Scaling up of highly active antiretroviral therapy in a rural district of Malawi: An effectiveness assessment. *Lancet*, *367*(9519), 1335–1342.

Figueroa, S. C., De Gatta, M. F., García, L. H., Hurlé, A. D.-G., Bernal, C. B., Correa, R. S., & Sánchez, M. J. G. (2010). The convergence of therapeutic drug monitoring and pharmacogenetic testing to optimize efavirenz therapy. *Therapeutic Drug Monitoring*, *32*(5), 579–585.

Folwaczny, M., Tengler, B., & Glas, J. (2012). Variants of the human NR1I2 (PXR) locus in chronic periodontitis. *Journal of Periodontal Research*, 47(2), 174-179.

Ford, N., Calmy, A., & Mofenson, L. (2011). Safety of efavirenz in the first trimester of pregnancy: An updated systematic review and meta-analysis. *AIDS*, *25*(18), 2301–2304.

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., & Dubchak, I. (2004). VISTA: Computational tools for comparative genomics. *Nucleic Acids Research*, *32*, W273–W279.

Gandhi, M., Ameli, N., Bacchetti, P., Gange, S. J., Anastos, K., Levine, A. M., … Greenblatt, R. M. (2009). Protease inhibitor levels in hair samples strongly predict virologic responses to HIV treatment. *AIDS*, *23*(4), 471–478.

Gandhi, M., Greenblatt, R. M., Bacchetti, P., Jin, C., Huang, Y., Anastos, K., … Aouizerat, B. (2012). A single-nucleotide polymorphism in *CYP2B6* leads to > 3-fold increases in efavirenz concentrations in plasma and hair among HIV-infected women. *The Journal of Infectious Diseases*, *206*(9), 1453–1461.

Gandhi, M., Mwesigwa, J., Aweeka, F., Plenty, A., Charlebois, E., Ruel, T. D., ... & Cohan, D. (2013). Hair and plasma data show that lopinavir, ritonavir, and efavirenz all transfer from mother to infant in utero, but only efavirenz transfers via breastfeeding. JAIDS Journal of Acquired Immune Deficiency Syndromes, 63(5), 578-584

Gatell, J. M., Katlama, C., Grinsztejn, B., Eron, J. J., Lazzarin, A., Vittecoq, D., … Nguyen, B.-Y. T. (2010). Long-term efficacy and safety of the HIV integrase inhibitor raltegravir in patients with limited treatment options in a Phase II study. *Journal of Acquired Immune Deficiency Syndromes*, *53*(4), 456–463.

Geick, A., Eichelbaum, M., & Burk, O. (2001). Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *The Journal of Biological Chemistry*, *276*(18), 14581–14587.

Gengiah, T. N., Holford, N. H. G., Botha, J. H., Gray, A. L., Naidoo, K., & Abdool Karim, S. S. (2012). The influence of tuberculosis treatment on efavirenz clearance in patients co-infected with HIV and tuberculosis. *European Journal of Clinical Pharmacology*, *68*(5), 689–695.

Giguère, V. (1999). Orphan nuclear receptors: From gene to function. *Endocrine Reviews*, *20*(5), 689–725.

Glas, J., Seiderer, J., Fischer, D., Tengler, B., Pfennig, S., Wetzke, M., … Brand, S. (2011). Pregnane X receptor (*PXR*/*NR1I2*) gene haplotypes modulate susceptibility to inflammatory bowel disease. *Inflammatory Bowel Diseases*, *17*(9), 1917–1924.

Grabe, N. (2002). AliBaba2: Context specific identification of transcription factor binding sites. *In Silico Biology*, *2*, S1–S15.

Gutiérrez, F., Navarro, A., Padilla, S., Antón, R., Masiá, M., Borrás, J., & Martín-Hidalgo, A. (2013). Prediction of neuropsychiatric adverse events associated with long-term efavirenz therapy, using plasma drug level monitoring. *Clinical Infectious Diseases*, *41*(11), 1648–1653.

Haas, D. W., Ribaudo, H. J., Kim, R. B., Tierney, C., Wilkinson, G. R., Gulick, R. M., … Acosta, E. P. (2004). Pharmacogenetics of efavirenz and central nervous system side effects: An adult AIDS clinical trials group study. *AIDS*, *18*(18), 2391–2400.

Haas, D. W., Smeaton, L. M., Shafer, R. W., Robbins, G. K., Morse, G. D., Labbe, L., … Kim, R. B. (2005). Pharmacogenetics of long-term responses to antiretroviral regimens containing efavirenz and/or nelfinavir: An adult AIDS clinical trials group study. *The Journal of Infectious Diseases*, *192*(11), 1931–1942.

Haas, D. W., Gebretsadik, T., Mayo, G., Menon, U. N., Acosta, E. P., Shintani, A., … Wilkinson, G. R. (2009). Associations between *CYP2B6* polymorphisms and pharmacokinetics after a single dose of nevirapine or efavirenz in African americans. *The Journal of Infectious Diseases*, *199*(6), 872–880.

Habtewold, Abiy, Amogne, W., Makonnen, E., Yimer, G., Riedel, K.-D., Ueda, N., … Aklillu, E. (2011). Long-term effect of efavirenz autoinduction on plasma/peripheral blood mononuclear cell drug exposure and CD4 count is influenced by *UGT2B7* and *CYP2B6* genotypes among HIV patients. *The Journal of Antimicrobial Chemotherapy*, *66*(10), 2350–2361.

Habtewold, A, Amogne, W., Makonnen, E., Yimer, G., Nylén, H., Riedel, K.-D., … Aklillu, E. (2012). Pharmacogenetic and pharmacokinetic aspects of *CYP3A* induction by efavirenz in HIV patients. *The Pharmacogenomics Journal*, *13*(6) 1–6.

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence lignment editor and analysis program for Windoes 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Hamers, R. L., Sigaloff, K. C. E., Wensing, A. M., Wallis, C. L., Kityo, C., Siwale, M., … Schuurman, R. (2012). Patterns of HIV-1 drug resistance after first-line antiretroviral therapy (ART) failure in 6 sub-Saharan African countries: Implications for second-line ART strategies. *Clinical Infectious Diseases*, *54*(11), 1660–1669.

Han, X. M., Ouyang, D. S., Chen, X. P., Shu, Y., Jiang, C. H., Tan, Z. R., & Zhou, H. H. (2002). Inducibility of CYP1A2 by omeprazole in vivo related to the genetic polymorphism of CYP1A2. *British Journal of Clinical Pharmacology*, *54*(5), 540-543.

Handschin, C., & Meyer, U. A. (2003). Induction of drug metabolism: The role of nuclear receptors. *Pharmacological Reviews*, *55*(4), 649–673.

Hariparsad, N., Nallani, S. C., Sane, R. S., Buckley, D. J., Buckley, A. R., & Desai, P. B. (2004). Induction of *CYP3A4* by efavirenz in primary human hepatocytes: Comparison with rifampin and phenobarbital. *Journal of Clinical Pharmacology*, *44*(11), 1273–1281.

Harries, A D, Nyangulu, D. S., Hargreaves, N. J., Kaluwa, O., & Salaniponi, F. M. (2001). Preventing antiretroviral anarchy in sub-Saharan Africa. *Lancet*, *358*, 410–414.

Harries, Anthony D, Zachariah, R., van Oosterhout, J. J., Reid, S. D., Hosseinipour, M. C., Arendt, V., … Kamoto, K. (2010). Diagnosis and management of antiretroviral-therapy failure in resource-limited settings in sub-Saharan Africa: Challenges and perspectives. *Lancet Infectious Diseases*, *10*(1), 60–65.

Hart, E., Curtis, H., Wilkins, E., & Johnson, M. (2007). National review of first treatment change after starting highly active antiretroviral therapy in antiretroviral-naïve patients. *HIV Medicine*, *8*(3), 186–191.

Hasegawa, M., Tahara, H., Inoue, R., Kakuni, M., Tateno, C., & Ushiki, J. (2012). Investigation of drug-drug interactions caused by human pregnane X receptor-mediated induction of CYP3A4 and CYP2C subfamilies in chimeric mice with a humanized liver. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *40*(3), 474–480.

Havlir, D. V, Bassett, R., Levitan, D., Gilbert, P., Tebas, P., Collier, A. C., … Wong, J. K. (2001). Prevalence and predictive value of intermittent viremia with combination HIV therapy. *Journal of the American Medical Association*, *286*(2), 171–179.

He, P., Court, M. H., Greenblatt, D. J., & Von Moltke, L. L. (2006). Human pregnane X receptor: genetic polymorphisms, alternative mRNA splice variants, and cytochrome P450 3A metabolic activity. *The Journal of Clinical Pharmacology*, *46*, 1356–1369.

Huang, Y., Gandhi, M., Greenblatt, R. M., Gee, W., Lin, E. T., & Messenkoff, N. (2008). Sensitive analysis of anti-HIV drugs, efavirenz, lopinavir and ritonavir, in human hair by liquid chromatography coupled with tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, *22*, 3401–3409.

Huang, L., Parikh, S., Rosenthal, P. J., Lizak, P., Marzan, F., Dorsey, G., … Aweeka, F. T. (2012). Concomitant efavirenz reduces pharmacokinetic exposure to the antimalarial drug artemether-lumefantrine in healthy volunteers. *Journal of Acquired Immune Deficiency Syndromes*, *61*(3), 310–316.

Hustert, E., Zibat, A., Presecan-siedel, E., Eiselt, R., Mueller, R., Fuß, C., … Ag, E. B. (2001). Natural protein variants of pregnane X receptor with altered transactivation activity toward *CYP3A4*. *Drug Metabolism and Disposition*, *29*(11), 1454–1459.

Ihunnah, C. A., Jiang, M., & Xie, W. (2011). Nuclear receptor PXR, transcriptional circuits and metabolic relevance. *Biochimica et Biophysica Acta*, *1812*(8), 956–963.

Ikeda, S., Kurose, K., Ozawa, S., Sai, K., Hasegawa, R., Komamura, K., … Minami, N. (2003). Twenty-six novel single nucleotide polymorphisms and their frequencies of the *NR1I3* (

CAR) gene in a Japanese population. *Drug Metabolism and Pharmacokinetics*, *18*(6), 413–418.

Ikeda, S., Kurose, K., Jinno, H., Sai, K., Ozawa, S., Hasegawa, R., … Sawada, J. (2005). Functional analysis of four naturally occurring variants of human constitutive androstane receptor. *Molecular Genetics and Metabolism*, *86*(1-2), 314–319.

Ikeda, S., Arai, T., Mieno, M., Tanaka, N., Sawabe, M. & Muramatsu, M. (2010). Association of genetic polymorphisms in human NR1I3 with hyperlipidemia in elderly Japanese population. HUGO Human Genome Meeting, Montpellier, France.

Ikediobi, O., Aouizerat, B., Xiao, Y., Gandhi, M., Gebhardt, S., & Warnich, L. (2011). Analysis of pharmacogenetic traits in two distinct South African populations. *Human Genomics*, *5*(4).

James, C., Preininger, L., & Sweet, M. (2012). Rilpivirine: A second-generation nonnucleoside reverse transcriptase inhibitor. *American Journal of Health-System Pharmacy*, *69*(10), 857–861.

Jeantils, V., Khuong, M., Delassus, J.-L., Honoré, P., Taverne, B., Uzan, M., & Tassi, S. (2006). [Efavirenz (Sustiva) in pregnancy: a study about 12 HIV patients]. *Gynécologie, Obstétrique & Fertilité*, *34*(7-8), 593–659.

Ji, H. Y., Lee, H., Lim, S. R., Kim, J. H., & Lee, H. S. (2012). Effect of efavirenz on UDP-glucuronosyltransferase 1A1, 1A4, 1A6, and 1A9 activities in human liver microsomes. *Molecules (Basel, Switzerland)*, *17*(1), 851–860.

Johnson, D. H., Gebretsadik, T., Shintani, A., Mayo, G., Acosta, E. P., Stein, C. M., & Haas, D. W. (2013). Neuropsychometric correlates of efavirenz pharmacokinetics and pharmacogenetics following a single oral dose. *British Journal of Cinical Pharmacology*, *75*(4), 997–1006.

Jones, S. A., Moore, L. B., Shenk, J. L., Wisely, G. B., Hamilton, G. A., Mckee, D. D., … Carolina, N. (2000). The pregnane Xreceptor: A promiscuous xenobiotic receptor that has diverged during evolution. *Molecular Endocrinology*, *14*, 27–39.

Katlama, C., Bonaventura, C., Mills, A., Trottier, B., Molina, J.-M., Grinsztejn, B., … Witek, J. (2010). Efficacy and safety of etravirine at week 96 in treatment-experienced HIV type-1-infected patients in the DUET-1 and DUET-2 trials. *Antiviral Therapy*, *15*(7), 1045–1052.

Kel, a. E. (2003). MATCHTM: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Research*, *31*(13), 3576–3579.

King, C. R., Xiao, M., Yu, J., Minton, M. R., Addleman, N. J., Van Booven, D. J., … Marsh, S. (2007). Identification of *NR1I2* genetic variation using resequencing. *European Journal of Clinical Pharmacology*, *63*(6), 547–554.

Klein, K., & Zanger, U. M. (2013). Pharmacogenomics of cytochrome P450 3A4: Recent progress toward the “missing heritability” problem. *Frontiers in Genetics*, *4*.

Kliewer, S. A, Moore, J. T., Wade, L., Staudinger, J. L., Watson, M. a, Jones, S. A, … Lehmann, J. M. (1998). An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. *Cell*, *92*(1), 73–82.

Kumarasamy, N., Vallabhaneni, S., Cecelia, A. J., Yepthomi, T., Balakrishnan, P., Saghayam, S., … Mayer, K. H. (2006). Reasons for modification of generic highly active antiretroviral therapeutic regimens among patients in southern India. *Journal of Acquired Immune Deficiency Syndromes*, *41*(1), 53–58.

Kumari, G., & Singh, R. K. (2012). Highly Active Antiretroviral Therapy for treatment of HIV/AIDS patients: Current status and future prospects and the Indian scenario. *HIV & AIDS Review*, *11*(1), 5–14.

Kwara, A., Lartey, M., Sagoe, K. W. C., Kenu, E., & Court, M. H. (2009). *CYP2B6*, CYP2A6 and UGT2B7 genetic polymorphisms are predictors of efavirenz mid-dose concentration in HIV-infected patients. *AIDS*, *23*(16), 2101–2106.

Lamba, V., Lamba, J., Yasuda, K., Strom, S., Davila, J., Hancock, M. L., … Schuetz, E. G. (2003). Hepatic *CYP2B6* expression: Gender and ethnic differences and relationship to *CYP2B6* genotype and CAR (constitutive androstane receptor) expression. *The Journal of Experimental Pharmacology and Therapeutics*, *307*(3), 906–922.

Lamba, J. K., Lamba, V., Yasuda, K., Lin, Y. S., Assem, M., Thompson, E., … Schuetz, E. (2004). Expression of constitutive androstane receptor splice variants in human tissues and their functional consequences. *The Journal of Pharmacology and Experimental Therapeutics*, *311*(2), 811–821.

Lamba, J. K. (2008). Pharmacogenetics of the constitutive androstane receptor. *Pharmacogenomics*, *9*(1), 71–83.

Lamba, J., Lamba, V., Strom, S., Venkataramanan, R., & Schuetz, E. (2008). Novel single nucleotide polymorphisms in the promoter and intron 1 of human pregnane X receptor/*NR1I2* and their association with *CYP3A4* expression. *Drug Metabolism and Disposition*, *36*(1), 169–181.

Laurent, C., Kouanfack, C., Laborde-Balen, G., Aghokeng, A. F., Mbougua, J. B. T., Boyer, S., … Delaporte, E. (2011). Monitoring of HIV viral loads, CD4 cell counts, and clinical assessments versus clinical monitoring alone for antiretroviral therapy in rural district hospitals in Cameroon (Stratall ANRS 12110/ESTHER): A randomised non-inferiority trial. *The Lancet Infectious Diseases*, *11*(11), 825–833.

Lazzarin, A., Clotet, B., Cooper, D., Reynes, J., Arasteh, K., Nelson, M., … Salgo, M. (2003). Efficacy of enfuvirtide in patients infected with drug-resistant HIV-1 in Europe and Australia. *The New England Journal of Medicine*, *348*(22), 2186–95.

LeCluyse, E. L. (2001). Pregnane X receptor: Molecular basis for species differences in *CYP3A* induction by xenobiotics. *Chemico-Biological Interactions*, *134*, 283–289.

Lee, W. M. (2003). Drug-Induced Hepatotoxicity. *The New England Journal of Medicine*, *349*(5), 474–485.

Lehmann, J. M., McKee, D. D., Watson, M. A, Willson, T. M., Moore, J. T., & Kliewer, S. A. (1998). The human orphan nuclear receptor PXR is activated by compounds that regulate *CYP3A4* gene expression and cause drug interactions. *The Journal of Clinical Investigation*, *102*(5), 1016–1023.

Lima, V. D., Nosyk, B., Wood, E., Kozai, T., Zhang, W., Chan, K., & Montaner, J. S. G. (2012). Assessing the effectiveness of antiretroviral regimens in cohort studies involving HIV-positive injection drug users. *AIDS*, *26*(12), 1491–1500.

Lindenberg, M., Kopp, S., & Dressman, J. B. (2004). Classification of orally administered drugs on the World Health Organization Model list of Essential Medicines according to the biopharmaceutics classification system. *European Journal of Pharmaceutics and Biopharmaceutics*, *58*(2), 265–278.

Livingston, R. J., von Niederhausern, A., Jegga, A. G., Crawford, D. C., Carlson, C. S., Rieder, M. J., … Nickerson, D. A. (2004). Pattern of sequence variation across 213 environmental response genes. *Genome Research*, *14*(10A), 1821–1831.

Loots, G. G., Ovcharenko, I., Pachter, L., Dubchak, I., & Rubin, E. M. (2002). rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Research*, *12*(5), 832–839.

Loutfy, M. R., Wu, W., Letchumanan, M., Bondy, L., Antoniou, T., Margolese, S., … Rochon, P. A. (2013). Systematic review of HIV transmission between heterosexual serodiscordant couples where the HIV-positive partner is fully suppressed on antiretroviral therapy. *PloS ONE*, *8*(2), e55747.

Lubomirov, R., Colombo, S., di Iulio, J., Ledergerber, B., Martinez, R., Cavassini, M., … Telenti, A. (2011). Association of pharmacogenetic markers with premature discontinuation of first-line anti-HIV therapy: An observational cohort study. *The Journal of Infectious Diseases*, *203*(2), 246–257.

Macallan, D. C. (1999). Nutrition and immune function in human immunodeficiency virus infection. *The Proceedings of the Nutrition Society*, *58*(3), 743–748.

Maglich, J. M., Sluder, A., Guan, X., Shi, Y., Mckee, D. D., Carrick, K., … Moore, J. T. (2001). Comparison of complete nuclear receptor sets from the human, Caenorhabditis elegans and Drosophila genomes. *Genome Biology*, *2*(8), 1–7.

Maglich, J. M., Stoltz, C. M., Goodwin, B., Hawkins-Brown, D., Moore, J. T., & Kliewer, S. A. (2002). Nuclear pregnane X receptor and constitutive androstane receptor regulate overlapping but distinct sets of genes involved in xenobiotic detoxification. *Molecular Pharmacology*, *62*(3), 638–646.

Maimbo, M., Kiyotani, K., Mushiroda, T., Masimirembwa, C., & Nakamura, Y. (2012). *CYP2B6* genotype is a strong predictor of systemic exposure to efavirenz in HIV-infected Zimbabweans. *European Journal of Clinical Pharmacology*, *68*(3), 267–271.

Mangelsdorf, D. J., & Evans, R. M. (1995). The RXR heterodimers and orphan receptors. *Cell*, *83*, 841–850.

Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., … Evans, R. M. (1995). The nuclear receptor superfamily: The second decade. *Cell*, *83*, 835–839.

Martín, M. T., Rovira, M., Massanes, M., del Cacho, E., Carcelero, E., Tuset, M., … Ribas, J. (2010). Analysis of the duration of and reasons for changing the first combination of antiretroviral therapy. *Farmacia Hospitalaria (English Edition)*, *34*(5), 224–230.

Martín, A. S., Figueroa, S. C., Guerrero, R. C., Hurtado, L. P., Hurlé, A. D.-G., & Álvarez, Á. C. (2013). Impact of pharmacogenetics on CNS side effects related to efavirenz. *Pharmacogenomics*, *14*(10), 1167–1178.

Marzolini, C., Telenti, A, Decosterd, L. A, Greub, G., Biollaz, J., & Buclin, T. (2001). Efavirenz plasma levels can predict treatment failure and central nervous system side effects in HIV-1-infected patients. *AIDS (London, England)*, *15*(1), 71–75.

Mckenna, N. J., & O'Malley, B. W. (2002). Nuclear receptor coactivators - An update. *Endocrinology*, *143*(7), 2461–2465.

Menezes, C. N., Duarte, R., Dickens, C., Dix-Peek, T., Van Amsterdam, D., John, M., … Crowther, N. J. (2013). The early effects of stavudine compared with tenofovir on adipocyte gene expression, mitochondrial DNA copy number and metabolic parameters in South African HIV-infected patients: a randomized trial. *HIV medicine*, *14*(4), 217–225.

Miller, C. M. P. (1997). Tools for Population Genetic Analyses (TFPGA): A Windows program for the analysis of allozyme and molecular population genetic data. Ver. 1.3. Retrieved from http://www.ccg.unam.mx/~vinuesa/tlem09/docs/TFPGADOC.PDF

Mollan, K., Daar, E. S., Sax, P. E., Balamane, M., Collier, A. C., Fischl, M. A, … Katzenstein, D. (2012). HIV-1 amino acid changes among participants with virologic failure: associations with first-line efavirenz or atazanavir plus ritonavir and disease status. *The Journal of Infectious Diseases*, *206*(12), 1920–1930.

Motsinger, A. A, Ritchie, M. D., Shafer, R. W., Robbins, G. K., Morse, G. D., Labbe, L., … Haas, D. W. (2006). Multilocus genetic interactions and response to efavirenz-containing regimens: An adult AIDS clinical trials group study. *Pharmacogenetics and Genomics*, *16*(11), 837–845.

Mukonzo, J. K., Röshammar, D., Waako, P., Andersson, M., Fukasawa, T., Milani, L., … Aklillu, E. (2009). A novel polymorphism in *ABCB1* gene, *CYP2B6**6 and sex predict single-dose efavirenz population pharmacokinetics in Ugandans. *British Journal of Clinical Pharmacology*, *68*(5), 690–699.

Mukonzo, J. K., Okwera, A., Nakasujja, N., Luzze, H., Sebuwufu, D., Ogwal-Okeng, J., … Aklillu, E. (2013). Influence of efavirenz pharmacokinetics and pharmacogenetics on neuropsychological disorders in Ugandan HIV-positive patients with or without tuberculosis: A prospective cohort study. *BMC Infectious Diseases*, *13*(1), 261.

Mutlib, A. E., Chen, H., Nemeth, G. A., Markwalder, J. A., Seitz, S. P., Gan, L. S., & Christ, D. D. (1999). Identification and characterization of efavirenz metabolites by liquid chromatography/mass spectrometry and high field NMR: Species difference in the metabolism of efavirenz. *Drug Metabolism and Disposition*, *27*(11), 1319–1333.

Mutwa, P. R., Fillekes, Q., Malgaz, M., Tuyishimire, D., Kraats, R. Van De Boer, K. R., … Geelen, S. P. M. (2012). Mid-dosing interval efavirenz plasma concentrations in HIV-1-infected children in Rwanda: treatment efficacy, tolerability, adherence, and the influence of *CYP2B6* polymorphisms. *Journal of Acquired Immune Deficiency Syndromes*, *60*(4), 400–404.

National Department of Health. (2004). *National Antiretroviral Treatment Guidelines* (pp. 1–50). Pretoria, South Africa.

National Department of Health. (2010). *The South African Antiretroviral Treatment Guidelines* (pp. 1–7). Pretoria, South Africa.

National Department of Health. (2013). *The South African Antiretroviral Treatment Guidelines* (pp. 1–21). Pretoria, South Africa.

Nevin, C. R., Ye, J., Aban, I., Mugavero, M. J., Jackson, D., Lin, H., … Willig, J. H. (2011). The role of toxicity-related regimen changes in the development of antiretroviral resistance. *AIDS Research and Human Retroviruses*, *27*, 1–7.

Ng, P. C. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812–3814.

Ngaimisi, E., Habtewold, A., Minzi, O., Makonnen, E., Mugusi, S., Mugusi, F., … Burhenne, J. (2013). Importance of ethnicity, *CYP2B6* and *ABCB1* genotype for efavirenz pharmacokinetics and treatment outcomes : A parallel-group prospective cohort study in two sub-Saharan Africa populations. *PLoS ONE*, *8*(7), e67946.

Nuclear Receptors Nomenclature Committee. (1999). A unified nomenclature system for the nuclear receptor superfamily. *Cell*, *97*, 161–163.

Osabe, M., Sugatani, J., Takemura, A., Kurosawa, M., Yamazaki, Y., Ikari, A., & Miwa, M. Up-regulation of CAR expression through Elk-1 in HepG2 and SW480 cells by serum starvation stress. *FEBS Letters*, *583*, 885-889.

Palella, F. J., Deloria-Knoll, M., Chmiel, J. S., Moorman, A. C., Wood, K. C., Greenberg, A. E., & Holmberg, S. D. (2003). Survival benefit of initiating antiretroviral therapy in HIV-infected persons in different CD4+ cell strata. *Annals of Internal Medicine*, *138*(8), 620-626.

Panel on Antiretroviral Guidelines for Adults and Adolescents. (2014). Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. *Department of Health and Human Services (DHHS)*, 1-285.

Pascussi, J. M., Dvorák, Z., Gerbal-Chaloin, S., Assenat, E., Maurel, P., & Vilarem, M. J. (2003). Pathophysiological factors affecting *CAR* gene expression. *Drug Metabolism Reviews*, *35*(4), 255–268.

Pascussi, J.M., Robert, A., Moreau, A., Ramos, J., Bioulac-Sage, P., Navarro, F., … Jose Vilarem, M. (2007). Differential regulation of constitutive androstane receptor expression by hepatocyte nuclear factor4α isoforms. *Hepatology*, *45*(5), 1146-1153.

Patterson, N., Petersen, D. C., van der Ross, R. E., Sudoyo, H., Glashoff, R. H., Marzuki, S., … Hayes, V. M. (2010). Genetic structure of a unique admixed population: Implications for medical research. *Human Molecular Genetics*, *19*(3), 411-419.

Pondugula, S. R., & Mani, S. (2013). Pregnane xenobiotic receptor in cancer pathogenesis and therapeutic response. *Cancer Letters*, *328*(1), 1–9.

Prosperi, M. C. F., Fabbiani, M., Fanti, I., Zaccarelli, M., Colafigli, M., Mondi, A., … Di Giambenedetto, S. (2012). Predictors of first-line antiretroviral therapy discontinuation due to drug-related adverse events in HIV-infected patients: a retrospective cohort study. *BMC Infectious Diseases*, *12*, 296.

Ramensky, V., Bork, P., & Sunyaev, S. (2002). Human non-synonymous SNPs: Server and survey. *Nucleic Acids Research*, *30*(17), 3894–3900.

Ramírez, J., Innocenti, F., Schuetz, E. G., Flockhart, D. A., Relling, M. V, Santucci, R., & Ratain, M. J. (2004). *CYP2B6*, *CYP3A4*, and *CYP2C19* are responsible for the *in vitro* N-demethylation of meperidine in human liver microsomes. *Drug Metabolism and Disposition*, *32*(9), 930–936.

Reschly, E. J., & Krasowski, M. D. (2006). Evolution and function of the NR1I nuclear hormone receptor subfamily (VDR, PXR, and CAR) with respect to metabolism of xenobiotics and endogenous compounds. *Current Drug Metabolism*, *7*(4), 349–365.

Richter, T., Mürdter, T. E., Heinkele, G., Pleiss, J., Tatzel, S., Schwab, M., … Zanger, U. M. (2004). Potent mechanism-based inhibition of human *CYP2B6* by clopidogrel and ticlopidine. *The Journal of Pharmacology and Experimental Therapeutics*, *308*(1), 189–197.

Ritchie, M. D., Haas, D. W., Motsinger, A. a, Donahue, J. P., Erdem, H., Raffanti, S., … Sterling, T. R. (2006). Drug transporter and metabolizing enzyme gene variants and nonnucleoside reverse-transcriptase inhibitor hepatotoxicity. *Clinical Infectious Diseases*, *43*(6), 779–82.

Rodríguez-Nóvoa, S., Barreiro, P., Jiménez-Nácher, I., & Soriano, V. (2006). Overview of the pharmacogenetics of HIV therapy. *The Pharmacogenomics Journal*, *6*(4), 234–245.

Röhrich, C. R. (2012). *The influence of pharmacogenetic traits and efavirenz levels on treatment outcome in HIV-positive South African women*. MSc Thesis. Stellenbosch University.

Rosenfeld, J. M., Vargas Jr, R., Xie, W., & Evans, R. M. (2003). Genetic profiling defines the xenobiotic gene network controlled by the nuclear receptor pregnane X receptor. *Molecular Endocrinology*, *17*(7), 1268–1282.

Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W., & Lenhard, B. (2004). JASPAR: An open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids research*, *32*(suppl 1), D91–D94.

Schapiro, J. M., Winters, M. A., Stewart, F., Efron, B., Norris, J., Kozal, M. J., & Merigan, T. C. (1996). The effect of high-dose saquinavir on viral load and CD4+ T-cell counts in HIV-infected patients. *Annals of Internal Medicine*, *48*(12), 1039–1050.

Schuster, S. C., Miller, W., Ratan, A., Tomsho, L. P., Giardine, B., Kasson, L. R., … Hayes, V. M. (2010). Complete Khoisan and Bantu genomes from southern Africa. *Nature*, *463*, 943–947.

Shubber, Z., Calmy, A., Andrieux-Meyer, I., Vitoria, M., Renaud-Thery, F., Shaffer, N., … Ford, N. (2013). Adverse events associated with nevirapine and efavirenz-based first-line antiretroviral therapy: A systematic review and meta-analysis. *AIDS*, *27*(9), 1403–1412.

Sinz, M., Kim, S., Zhu, Z., Chen, T., Anthony, M., Dickinson, K., & Rodrigues, A. (2006). Evaluation of 170 xenobiotics as transactivators of human pregnane X receptor (hPXR) and correlation to known CYP3A4 drug interactions. *Current Drug Metabolism*, *7*(4), 375–388.

Sridar, C., Kenaan, C., & Hollenberg, P. F. (2012). Inhibition of bupropion metabolism by selegiline: Mechanism-based inactivation of human *CYP2B6* and characterization of glutathione and peptide adducts. *Drug Metabolism and Disposition*, *40*(12), 2256–266.

Ståhle, L., Moberg, L., Svensson, J., & Sönnerborg, A. (2004). Efavirenz plasma concentrations in HIV-infected patients. *Therapeutic Drug Monitoring*, *26*(3), 267–270.

Statistics South Africa. (2004). *Provincial Profile: Western Cape* (pp. 1–91). Pretoria, South Africa.

Statistics South Africa. (2012). *Census 2011 in brief*. Pretoria, South Africa

Statistics South Africa. (2013). *Statistical Release P0302: Mid-year population estimates* (pp. 1–16). Pretoria, South Africa.

Steele, K. T., Steenhoff, A. P., Newcomb, C. W., Rantleru, T., Nthobatsang, R., Lesetedi, G., … Bisson, G. P. (2011). Early mortality and AIDS progression despite high initial antiretroviral therapy adherence and virologic suppression in Botswana. *PloS ONE*, *6*(6), e20010.

Stover, J., Gopalappa, C., Mahy, M., Doherty, M. C., Easterbrook, P. J., Weiler, G., & Ghys, P. D. (2014). The impact and cost of the 2013 WHO recommendations on eligibility for antiretroviral therapy. *AIDS*, *28*, S225-S230.

Sueyoshi, T, Kawamoto, T., Zelko, I., Honkakoski, P., & Negishi, M. (1999). The repressed nuclear receptor CAR responds to phenobarbital in activating the human *CYP2B6* gene. *The Journal of Biological Chemistry*, *274*(10), 6043–6046.

Sueyoshi, Tatsuya, Green, W. D., Vinal, K., Woodrum, T. S., Moore, R., & Negishi, M. (2011). Garlic extract diallyl sulfide (DAS) activates nuclear receptor CAR to induce the *Sult1e1* gene in mouse liver. *PloS ONE*, *6*(6), e21229.

Sugatani, J., Nishitani, S., Yamakawa, K., Yoshinari, K., Sueyoshi, T., Negishi, M., & Miwa, M. (2005). Transcriptional regulation of human *UGT1A1* gene expression: Activated glucocorticoid receptor enhances constitutive androstane receptor/pregnane X receptor-mediated UDP-glucuronosyltransferase 1A1 regulation with glucocorticoid receptor-interacting prote. *Molecular Pharmacology*, *67*(3), 845–855.

Sukasem, C., Chamnanphon, M., Koomdee, N., Puangpetch, A., Santon, S., Jantararoungtong, T., … Manosuthi, W. (2013). High plasma efavirenz concentration and *CYP2B6*

polymorphisms in Thai HIV-1 infections. *Drug Metabolism and Pharmacokinetics*, (66), 1–24.

Svärd, J., Spiers, J. P., Mulcahy, F., & Hennessy, M. (2010). Nuclear receptor-mediated induction of CYP450 by antiretrovirals: Functional consequences of *NR1I2* (PXR) polymorphisms and differential prevalence in whites and sub-Saharan Africans. *Journal of Acquired Immune Deficiency Syndromes*, *55*(5), 536–549.

Swart, M., Whitehorn, H., Ren, Y., Smith, P., Ramesar, R.S., & Dandara, C. (2012). PXR and CAR single nucleotide polymorphisms influence plasma efavirenz levels in South African HIV/AIDS patients. *BMC Medical Genetics*, 13(1), 112.

Swart, M., Skelton, M., Ren, Y., Smith, P., Takuva, S., & Dandara, C. (2013). High predictive value of *CYP2B6* SNPs for steady-state plasma efavirenz levels in South African HIV/AIDS patients. *Pharmacogenetics and Genomics*, *23*(8), 415–427.

Tabone, T., Mather, D. E., & Hayden, M. J. (2009). Temperature switch PCR (TSP): Robust assay design for reliable amplification and genotyping of SNPs. *BMC genomics*, *10*(1), 580.

Tashima, K. T., Caliendo, a M., Ahmad, M., Goömley, J. M., Fiske, W. D., Brennan, J. M., & Flanigan, T. P. (1999). Cerebrospinal fluid human immunodeficiency virus type 1 (HIV-1) suppression and efavirenz drug concentrations in HIV-1-infected patients receiving combination therapy. *The Journal of Infectious Diseases*, *180*(3), 862–864.

Taylor, G. P., Clayden, P., Dhar, J., Gandhi, K., Gilleece, Y., Harding, K., … De Ruiter, A. (2012). British HIV Association guidelines for the management of HIV infection in pregnant women. *HIV Medicine*, *13*(S2), 87-157

The 1000 Genomes Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*, 56–65.

The International HapMap 3 Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*(7311), 52–58.

The R Development Core Team. (2010). R : A language and environment for statistical computing. Retrieved from http://www.lsw.uni-heidelberg.de/users/christlieb/teaching/UKStaSS10/R-refman.pdf

Thiam, M., Diop-Ndiaye, H., Diouf, A. D., Vidal, N., Ndiaye, O., Ndiaye, I., … Toure-Kane, C. (2013). HIV-1 genetic diversity and drug resistance among Senegalese patients in the public health system. *Journal of Clinical Microbiology*, *51*(2), 578–584.

Thompson, E. E., Kuttab-Boulos, H., Krasowski, M. D., & Di Rienzo, A. (2005). Functional constraints on the constitutive androstane receptor inferred from human sequence variation and cross-species comparisons. *Human Genomics*, *2*(3), 168–178.

Thompson, M. A., Aberg, J. A., Hoy, J. F., Benson, C., Gu, H. F., Hammer, S. M., … Volberding, P. A. (2012). Antiretroviral treatment of adult HIV infection: 2012 recommendations of the International Antiviral Society - USA Panel. *Journal of the American Medical Association*, *308*(4), 387–402.

Tishkoff, S. A, Reed, F. A, Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., … Williams, S. M. (2009). The genetic structure and history of Africans and African Americans. *Science*, *324*(5930), 1035–1044.

Tojima, H., Kakizaki, S., Yamazaki, Y., Takizawa, D., Horiguchi, N., Sato, K., & Mori, M. (2012). Ligand dependent hepatic gene expression profiles of nuclear receptors CAR and PXR. *Toxicology Letters*, *212*(3), 288–297.

Tolson, A. H., Li, H., Eddington, N. D., & Wang, H. (2009). Methadone induces the expression of hepatic drug-metabolizing enzymes through the activation of pregnane X receptor and constitutive androstane receptor. *Drug Metabolism and Disposition*, *37*(9), 1887–1894.

Tsuchiya, K., Gatanaga, H., Tachikawa, N., Teruya, K., Kikuchi, Y., Yoshino, M., … Oka, S. (2004). Homozygous *CYP2B6* *6 (Q172H and K262R) correlates with high plasma efavirenz concentrations in HIV-1 patients treated with standard efavirenz-containing regimens. *Biochemical and Biophysical Research Communications*, *319*(4), 1322–1326.

United Nations. (1995). The Copenhagen Declaration and Programme of Action of the World Summit for Social Development (pp. 1–21). New York, USA.

Urano, T., Usui, T., Shiraki, M., Ouchi, Y., & Inoue, S. Association of a single nucleotide polymorphism in the constituive androstane receptor gene with bone mineral density. *Geriatrics & Gerontology International, 9,* 235-241.

Vacca, M., Degirolamo, C., Massafra, V., Polimeno, L., Mariani-Costantini, R., Palasciano, G., & Moschetta, A. (2013). Nuclear receptors in regenerating liver and hepatocellular carcinoma. *Molecular and Cellular Endocrinology*, 368(1), 108-119.

Victor, G.M. (2011). *The effect of functional SNPs in pharmacogenes on the levels of efavirenz in hair of two populations from South Africa.* Honours Dissertation. Stellenbosch University.

Wang, H., Faucette, S., Sueyoshi, T., Moore, R., Ferguson, S., Negishi, M., & LeCluyse, E. L. (2003). A novel distal enhancer module regulated by pregnane X receptor/constitutive androstane receptor is essential for the maximal induction of *CYP2B6* gene expression. *The Journal of Biological Chemistry*, *278*(16), 14146–14152.

Wang, H., & Lecluyse, E. L. (2003). Role of orphan nuclear receptors in the regulation of drug-metabolising enzymes. *Clinical Pharmacokinetics*, *42*(15), 1331–1357.

Wang, J., Sönnerborg, A., Rane, A., Josephson, F., Lundgren, S., Ståhle, L., & Ingelman-Sundberg, M. (2006). Identification of a novel specific *CYP2B6* allele in Africans causing impaired metabolism of the HIV drug efavirenz. *Pharmacogenetics and Genomics*, *16*(3), 191–198.

Ward, B. A., Gorski, J. C., Jones, D. R., Hall, S. D., & Flockhart, D. A. (2003). The cytochrome P450 2B6 (*CYP2B6*) is the main catalyst of efavirenz primary and secondary metabolism : implication for HIV / AIDS therapy and utility of efavirenz as a substrate marker of *CYP2B6* catalytic activity. *The Journal of Pharmacology and Experimental Therapeutics*, *306*(1), 287–300.

Wester, C. W., Kim, S., Bussmann, H., Avalos, A., Ndwapi, N., Peter, T. F., … Marlink, R. (2005). Initial response to highly active antiretroviral therapy in HIV-1C-infected adults in a public sector treatment program in Botswana. *Journal of Acquired Immune Deficiency Syndromes*, *40*(3), 336–343.

WHO Collaborating Centre for Drug Statistics Methodology. (2011). *Guidelines for ATC classification and DDD assignment* (pp. 1–286). Geneva, Switzerland.

Wojnowski, L. (2004). Genetics of the variable expression of CYP3A in humans. *Therapeutic Drug Monitoring*, *26*(2), 192-199.

Woldemedhin, B., & Wabe, N. T. (2012). The reason for regimen change among HIV/AIDS patients initiated on first line Highly Active Antiretroviral Therapy in Southern Ethiopia. *North American Journal of Medical Sciences*, *4*(1), 19–23.

World Health Organization. (2000). *Obesity: Preventing and managing the global epidemic* (pp. 1 – 253). Geneva, Switzerland.

World Health Organization. (2004). *Scaling up antiretroviral therapy in resource-limited settings: Treatment guidelines for a public health approach* (pp. 1–68). Geneva, Switzerland.

World Health Organization. (2006). *Antiretroviral therapy for HIV infection in adults and adolescents: Recommendations for a public health approach* (pp. 1–134). Geneva, Switzerland.

World Health Organization. (2010a). *Antiretroviral therapy for HIV infection in adults and adolescents: Recommendations for a public health approach* (pp. 1–156). Geneva, Switzerland.

World Health Organization. (2010b). *Towards universal access: Scaling up priority HIV/AIDS interventions in the health sector - Progress Report 2010*. Geneva, Switzerland.

World Health Organization. (2013a). *Global update on HIV treatment 2013: Results, impact and opportunities* (pp. 1–126). Geneva, Switzerland.

World Health Organization. (2013b). *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: Recommendations for a public health approach* (pp. 1–272). Geneva, Switzerland.

World Health Organization. (2013c). WHO Global Database on Body Mass Index. Retrieved December 10, 2013, from http://apps.who.int/bmi/index.jsp

World Health Organization, & UNAIDS. (2003). *AIDS epidemic update* (pp. 1–17). Geneva,Switzerland

World Health Organization. (2014). *Global update on the health sector response to HIV 2014.* (pp. 1–168). Geneva, Switzerland.

World Health Organization, & UNAIDS. (2006). *Progress on global access to HIV antiretroviral therapy: A report on "3 by 5" and beyond* (pp. 1–81). Geneva, Switzerland.

Wyen, C., Hendra, H., Vogel, M., Hoffmann, C., Knechten, H., Brockmeyer, N. H., … Owen, A. (2008). Impact of *CYP2B6* 983T>C polymorphism on non-nucleoside reverse transcriptase inhibitor plasma concentrations in HIV-infected patients. *The Journal of Antimicrobial Chemotherapy*, *61*(4), 914–918.

Wyen, C., Hendra, H., Siccardi, M., Platten, M., Jaeger, H., Harrer, T., … Owen, A. (2011). Cytochrome P450 2B6 (*CYP2B6*) and constitutive androstane receptor (CAR) polymorphisms are associated with early discontinuation of efavirenz-containing regimens. *The Journal of Antimicrobial Chemotherapy*, *6*, 1–7.

Xiao, L., Xie, X., & Zhai, Y. (2010). Functional crosstalk of CAR-LXR and ROR-LXR in drug metabolism and lipid metabolism. *Advanced Drug Delivery Reviews*, *62*(13), 1316–1321.

Xie, W., Radominska-Pandya, A., Shi, Y., Simon, C. M., Nelson, M. C., Ong, E. S., … Evans, R. M. (2001). An essential role for nuclear receptors SXR/PXR in detoxification of cholestatic bile acids. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(6), 3375–3380.

Xie, W., Yeuh, M.-F., Radominska-Pandya, A., Saini, S. P. S., Negishi, Y., Bottroff, B. S., … Evans, R. M. (2002). Control of steroid, heme , and carcinogen metabolism by nuclear pregnane X receptor and constitutive androstane receptor. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(7), 4150–4155.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*, 134.

Yimer, G., Ueda, N., Habtewold, A., Amogne, W., Suda, A., Riedel, K.-D., … Aklillu, E. (2011). Pharmacogenetic & pharmacokinetic biomarker for efavirenz based ARV and rifampicin based anti-TB drug induced liver injury in TB-HIV infected patients. *PloS ONE*, *6*(12), e27810.

Yoshinari, K., Yoda, N., Toriyabe, T., & Yamazoe, Y. (2010). Constitutive androstane receptor transcriptionally activates human *CYP1A1* and *CYP1A2* genes through a common regulatory element in the 5'-flanking region. *Biochemical Pharmacology*, *79*(2), 261–269.

Zanger, U. M., Turpeinen, M., Klein, K., & Schwab, M. (2008). Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. *Analytical and Bioanalytical Chemistry*, *392*(6), 1093–108.

Zanger, U. M., & Schwab, M. (2013). Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & Therapeutics*, *138*(1), 103–141.

Zhang, J., Kuehl, P., Green, E. D., Touchman, J. W., Hudson, J. R., Schuetz, E. G., & Boguski, M. S. (2001). The human pregnane X receptor: Genomic structure and identification and functional characterization of natural allelic variants. *Pharmacogenetics*, *11*, 555–572.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406–3415.

Electronic References

Ensembl Genomes Browser. http://www.ensembl.org/index.html

Google Scholar: http://scholar.google.co.za/schhp?hl=en&as_sdt=0,5

HapMap Genomes Browser: http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap28_B36/

HomoloGene: http://www.ncbi.nlm.nih.gov/homologene

Pubmed. http://www.ncbi.nlm.nih.gov/pubmed

ScienceDirect. http://www.sciencedirect.com/

The 1000 Genomes Browser. http://browser.1000genomes.org/index.html

VassarStats: Website for Statistical Computation. http://www.vassarstats.net/

Appendices

Appendix A: Protocols

Laboratory Protocol for Manual Purification of DNA from 0.5 ml of Oragene®•DNA/saliva (DNA Genotek Inc., Kanata, Canada).

DNA yield and stability in Oragene·DNA

Oragene•DNA is designed to capture the large amount of DNA present in saliva. The median yield of DNA from 2 ml of saliva when captured in 2 ml of Oragene•DNA is 110 µg.

The following step-by-step protocol describes how to purify DNA from a 500 µl aliquot of an Oragene•DNA/saliva sample. Volumes of less than 500 µl may be purified by adjusting the volumes of reagents accordingly.

Equipment and reagents to be supplied by user

- Microcentrifuge capable of running at 13,000 rpm (15,000 × g)
- Air or water incubator at 50°C (Note: water incubator is not recommended for OG-500).
- Ethanol (95 to 100%) at room temperature
- DNA buffer: TE (10 mM Tris-HCl, 1mM EDTA, pH 8.0) or similar solution
- (Optional) Glycogen (20 mg/ml) (e.g., Invitrogen Cat. No. 10814-010)
- Ethanol (70%) at room temperature

Procedure

Purification Steps	Notes
1. Mix the Oragene•DNA/saliva sample in the Oragene•DNA vial by inversion and gentle shaking for a few seconds.	• This is to ensure that viscous saliva samples are properly mixed with the Oragene•DNA solution.
2. Incubate the sample at 50°C in a water incubator for a minimum of one hour or in an air incubator for a minimum of two hours.	• DNA in Oragene•DNA is stable at room temperature even without the incubation step. • This heat-treatment step is essential to ensure that DNA is adequately released and that nucleases are permanently inactivated. • This incubation step may be performed at any time after saliva is collected and before it is purified. • Incubation of the entire sample is recommended to ensure that this step has been completed. However, if it is more convenient, a 500 µl aliquot can be heat-treated. • The sample can be incubated either in the original container or after transfer to another tube. • The sample may be incubated at 50°C overnight if it is more convenient. • A longer time is required in an air incubator because temperature equilibration is slower than in a water incubator.
3. Transfer 500 µl of the mixed Oragene•DNA/saliva sample to a 1.5 ml microcentrifuge tube.	• The remainder of the Oragene•DNA/saliva sample can be stored at room temperature or frozen (-15°C to -20°C).
4. For 500 µl of Oragene•DNA/saliva, add 20 µl	• The sample will become turbid as impurities and

(1/25th volume) of Oragene•DNA Purifier (OG-L2P, supplied) to the microcentrifuge tube and mix by vortexing for a few seconds.	inhibitors are precipitated.
5. Incubate on ice for 10 minutes.	• Room temperature incubation can be substituted but will be slightly less effective in removing impurities.
6. Centrifuge at room temperature for five minutes at 13,000 rpm (15,000 × g).	• A longer period of centrifugation (up to 15 min) may be beneficial in reducing the turbidity (high A320) of the final DNA solution.
7. Carefully transfer the clear supernatant with a pipet tip into a fresh microcentrifuge tube. Discard the pellet containing impurities.	• The pellet contains turbid impurities. If accidentally disturbed, the tube should be re-centrifuged.
Optional: Addition of Glycogen	Optional: Addition of Glycogen Some users may prefer to add 5 µl (100 µg) of Glycogen to the supernatant to make the pellet more easily visible.
8. To 500 µl of supernatant, add 500 µl (i.e., an equal volume) of room-temperature 95-100% ethanol. Mix gently by inversion 10 times.	• During mixing with ethanol, the DNA will be precipitated. This may appear as a clot of DNA fibers or as a fi ne precipitate, depending upon the amount of DNA in the sample. • Even if no clot is seen, DNA will be recovered by carefully following the next steps.
9. Allow the sample to stand at room temperature for 10 minutes to allow the DNA to fully precipitate.	• Incubation at -20°C is not recommended because impurities may co-precipitate with the DNA.
10. Place the tube in the microcentrifuge in a known orientation. Centrifuge at room temperature for two minutes at 13,000 rpm (15,000 × g).	• For example, place each tube in the microcentrifuge with the hinge portion of the cap pointing away from the centre of the rotor. After centrifugation, the position of the pellet can be located (even if too tiny to be easily visible.): it will be at the tip of the tube below the hinge.
11. Carefully remove the supernatant with a pipet tip and discard it. Take care to avoid disturbing the DNA pellet.	• This pellet contains DNA. Loss of the pellet will result in loss of the DNA. • Rotating the tube such that the pellet is on the upper wall will allow you to safely move a pipet tip along the lower wall and remove all of the supernatant. • The supernatant may contain impurities and should be removed as completely as possible. • Excessive drying of the pellet can make the DNA more difficult to dissolve.
12. Ethanol wash step. Carefully add 250 µl of 70% ethanol. Let stand at room temperature for 1 minute. Completely remove the ethanol without disturbing the pellet.	• Take care not to disturb the DNA pellet. • The DNA pellet may be small. Addition of a carrier such as glycogen at step #7 will increase the visibility of the pellet. • Should the pellet detach, cetrifuge the sample for five minutes at 13,000 rpm (15,000 x g). • The 70% ethanol wash helps to remove residual • inhibitors.
13. Add 100 µl of DNA buffer to dissolve the DNA pellet. Vortex for at least five seconds.	• The expected concentration of the fully hydrated DNA is 20 to 200 ng/µl. • Note that large amounts of high molecular weight DNA can be slow to hydrate (dissolve) completely. • Incomplete hydration of the DNA is a cause of inaccuracy in estimating DNA concentration and of failure of downstream applications such as PCR.
14. (Optional) Additional steps to ensure complete hydration of the DNA.	a) Additional vigorous pipetting and vortexing, and/or b) Incubation at 50°C for 1 hour with occasional vortexing, and/or c) Incubation at room temperature for one to two days • For applications such as Southern blotting that require

	very high molecular weight DNA, (c) is recommended.
15. Storage of the fully rehydrated DNA.	• In TE at 4°C for up to 1-2 months. • Recommended in TE in aliquots at -20°C for long-term storage. • Note that freezing of purified DNA in TE will cause DNA to precipitate. When thawing a sample of frozen purified DNA, pay careful attention to rehydration, as discussed in step 13.

Bioline SureClean protocol for PCR clean-up

1. Add an equal volume of SureClean to nucleic acid solution and mix thoroughly.
2. Incubate at room temperature for at least 10 min.
3. Centrifuge at maximum speed (best results at 14,000x g) in a bench-top centrifuge for 10 minutes and carefully remove supernatant by aspiration.

(Note: Centrifuging for longer time leads to better DNA recovery)

4. Add a volume of 70% Ethanol equal to 2x original sample volume and vortex for 10 seconds.

(Note: For sensitive applications an optional second ethanol wash can be performed)

5. Centrifuge at maximum speed (best results at 14,000 x g) in a bench-top centrifuge for 10 minutes, remove supernatant and air-dry to ensure complete removal of ethanol. (Note: Do not overdry the pellet)

Resuspend pellet in desired volume of TE, water or any other appropriate buffer for downstream procedures.

Notes:

A. Apparent molecular weight of the DNA treated (agarose gel electrophoresis) may be higher if the washing-step with 70% ethanol step is omitted. For accurate MW assay, two washing steps are recommended after the cleaning procedure.

B. Nucleic acids to be purified must be ≥ 100 bp.

C. Easier visualisation of the pellet can be achieved with a pink co-precipitant which is supplied in the Bioline SureClean Plus (BIO-37047).

Appendix B: List of primers

Table B-1: Primers for bi-directional sequencing of *NR1I3*

Start of amplicon	Amplicon name	Amplicon length (bp)	Primer Name	Sequence (5'-3')	Size (bp)	T_M (°C)	GC (%)	Source
4791 bp upstream from exon 1	UA1	576	*NR1I3* Upstream Amplicon 1 F	ATTAGGCTCACAGAGGTGATCTGC	24	58.8	50.0	PrimerQuest
			NR1I3 Upstream Amplicon 1 R	CTCCCTCACTCCCTATAAATGACC	24	56.2	50.0	PrimerQuest
1245 bp upstream from exon 1	UA2	611	*NR1I3* Upstream Amplicon 2 F	GCTATTATTACAGGTGCACGCC	22	56.1	50.0	PrimerQuest
			NR1I3 Upstream Amplicon 2 R	AGAGCCCTCTCACTTCCTAAC	21	55.8	52.4	PrimerQuest
695 bp upstream from exon 1	UA3	791	*NR1I3* Upstream Amplicon 3 F	TGCCCTGTGGTCACTCTGATTT	22	58.7	50.0	PrimerQuest
			NR1I3 Upstream Amplicon 3 R	TTATGTGGCCTCCAGTTGCTCT	22	58.5	50.0	PrimerQuest
142 bp upstream from exon 1	Exon 1	520	*NR1I3* Exon 1 F primer	TCCCAGCTTGTTCAGGATCT	20	55.8	50.0	Thompson *et al.*, 2005
			NR1I3 Exon 1 R primer	AAGTGCTGGAATGACACACG	20	55.7	50.0	Thompson *et al.*, 2005
338 bp upstream from exon 2	Exon 2	636	*NR1I3* Exon 2 altF primer	AACCTGGGCAACATGGCAAA	20	58.3	50.0	PrimerQuest
			NR1I3 Exon 2 R primer	TGCCAGCCACAGGGTAGTTA	20	58.4	55.0	PrimerQuest
138 bp upstream from exon 3	Exon 3	379	*NR1I3* Exon 3 F primer	GCTTGAGATGTGTTGGATAC	20	51.0	45.0	Ikeda *et al.*, 2003
			NR1I3 Exon 3 R primer	TGTGCCTGCTATGTAGAGAA	20	53.1	45.0	Ikeda *et al.*, 2003
138 bp upstream from exon 4	Exon 4/5	814	*NR1I3* Exon 4/5 F primer	AGCCAAGCCCAAGAAAGTCT	20	56.9	50.0	Thompson *et al.*, 2005
			NR1I3 Exon 4/5 R primer	CCAAAGCTGGGAATTCAAGA	20	53.0	45.0	Thompson *et al.*, 2005
9 bp upstream from exon 6	Exon 6/7	519	*NR1I3* Exon 6/7 altF primer	AAGTGTCCTTGTTTCCTGC	19	53.0	47.4	PrimerQuest
			NR1I3 Exon 6/7 R primer	GACTGAGGCTGGGATAGACCT	21	57.9	57.1	PrimerQuest
64 bp upstream from exon 8	Exon 8	317	*NR1I3* Exon 8 F primer	GTCTATCCCAGCCTCAGTCC	22	56.7	60.0	Ikeda *et al.*, 2003
			NR1I3 Exon 8 R primer	TAGACCCCACGATACCTGAA	20	54.7	50.0	Ikeda *et al.*, 2003
175 bp upstream from exon 9	Exon 9	556	*NR1I3* Exon 9 F primer	AGACAGTGTCAGACCGAGGC	20	59.0	60.0	Thompson *et al.*, 2005
			NR1I3 Exon 9 R primer	TTAGCCCTGAGGTTTCCTCC	20	56.4	55.0	Thompson *et al.*, 2005

Table B-2: Primers for genotyping of *NR1I3* SNPs

Position	SNP	Genotyping Assay	Amplicon length (bp)	Primer Name	Sequence (5'-3')	Size (bp)	T_M (°C)	GC (%)	Source
Exon 7	rs113800307	PCR-RFLP	519	*NR1I3* Exon 6/7 altF primer	AAGTGTCCTTGTTTCCTGC	19	53.0	47.4	PrimerQuest
				NR1I3 Exon 6/7 R primer	GACTGAGGCTGGGATAGACCT	21	57.9	57.1	PrimerQuest
Intron 5	rs2307420	PCR-RFLP	277	CAR rs2307420 F	CCTCATCTTTCAGGAGGCAAACACT**g**C	27	61.5	51.9	PrimerQuest
				CAR rs2307420 R	CAAACTTCTGGACTCAAGCGATCTGCC	27	61.4	51.9	PrimerQuest
Exon 5	rs2307424	TSP	856	*NR1I3* TSP LS 1 F	TGCAGTTCTTCACAGGTTGGTGGA	24	60.7	50.0	PrimerQuest
				NR1I3 TSP LS 1 R	ATGCTTGGAAAGGCTGACGCATGT	24	61.7	50.0	PrimerQuest
			218 / 324	*NR1I3* rs2307424 NLS F	GCCCAAGATAAGGCTGG	17	53.0	58.8	PrimerQuest
				NR1I3 rs2307424 NLS R	CGTCACCGGAAGACG	15	52.7	66.7	PrimerQuest
Exon 4	rs34161743	PCR-RFLP	814	*NR1I3* Exon 4/5 F primer	AGCCAAGCCCAAGAAAGTCT	20	56.9	50.0	Thompson *et al.*, 2005
				NR1I3 Exon 4/5 R primer	CCAAAGCTGGGAATTCAAGA	20	53.0	45.0	Thompson *et al.*, 2005
Exon 4	rs35205211	PCR-RFLP	814	*NR1I3* Exon 4/5 F primer	AGCCAAGCCCAAGAAAGTCT	20	56.9	50.0	Thompson *et al.*, 2005
				NR1I3 Exon 4/5 R primer	CCAAAGCTGGGAATTCAAGA	20	53.0	45.0	Thompson *et al.*, 2005
Intron 3	rs2502815	TSP	733	*NR1I3* TSP LS 4 F	TTGCAGTGAGCCGAGATTGCTCCATT	24	62.6	50.0	PrimerQuest
				NR1I3 TSP LS 4 R	TGTCATGATCTCTTGCCTGGCACA	24	60.6	50.0	PrimerQuest
			229 / 315	*NR1I3* rs2502815 NLS F	GCCCTCCCTCTACACTG	17	54.4	64.7	PrimerQuest
				NR1I3 rs2502815 NLS R	CGCCCTGCACAACG	14	54.6	71.4	PrimerQuest
Intron 4	rs3003596[1]	PCR-RFLP	204	CAR rs3003596 F	CTTGCTTTTTCCATATCACACGA	23	62.0	39.1	PrimerQuest
				CAR rs3003596 R	AGAAATGTGGACAGGCTTAGAT	22	62.0	40.9	PrimerQuest
Intron 2	rs35709078	PCR-RFLP	379	*NR1I3* Exon 3 F primer	GCTTGAGATGTGTTGGATAC	20	51.0	45.0	Ikeda *et al.*, 2003
				NR1I3 Exon 3 R primer	TGTGCCTGCTATGTAGAGAA	20	53.1	45.0	Ikeda *et al.*, 2003
5'-upstream	rs55802895	PCR-RFLP	538	*NR1I3* TSP LS 3 F	AAATGACAGGTGCAGCCCTTGACT	24	61.1	50.0	PrimerQuest
				NR1I3 TSP LS 3 R	TTCCAACCCATCCTCCATGCTGTT	24	60.8	50.0	PrimerQuest

[1]Primers were adapted from Swart *et al.*, 2012. Bold lowercase letters are incorporated mutagenic nucleotides. Underlined letters are non-complementary 5'-tails.

Table B-3: Primers for genotyping of *NR1I2* SNPs

Position	SNP	Genotyping Assay	Amplicon length (bp)	Primer Name	Sequence (5'-3')	Size (bp)	T_M (°C)	GC (%)	Source
5'-upstream	rs3814055	PCR-RFLP	504	PXR TSP LS 1 F	TCCCAGCCTCAAGTCTTTCTCATCCCTT	28	62.7	50	PrimerQuest
				PXR TSP LS 1 R	AGTCCTCACTCTTCAGCCTCAACCTTCT	28	62.3	50	PrimerQuest
			221	PXR rs3814055 F	TTTTTTGGCAATCCCAGG**a**T	20	52.8	40	Andrews *et al.*, 2010
				PXR rs3814055 R	CGAATGTGGTGGATACCAG	19	52.9	52.6	Andrews *et al.*, 2010
5'-upstream	rs1523128	PCR-RFLP	696	PXR rs1523128 F	TCCTGGTCAGCCTTCTGTTCTGAG	24	60.3	54.2	PrimerQuest
				PXR rs1523128 R	AGGAGTAGCTTTACCAAGGAGCCG	24	60.3	54.2	PrimerQuest
5'-UTR	rs1523127	TSP	450	PXR TSP LS 2 F	AGCCGCTTAGTGCCTACATCTGACTT	26	61.6	50	PrimerQuest
				PXR TSP LS 2 R	AGCAGCAAGGCCAACATGATTCACGA	26	63	50	PrimerQuest
			233 / 420	PXR rs1523127 NLS F	GCTACTCCTGTCCTGAAC	18	52	55.6	PrimerQuest
				PXR rs1523127 NLS R	CGCTATGATGCTGACCTC	18	52.8	55.6	PrimerQuest
5'-UTR	rs3842689	PCR-RFLP	112 / 106	PXR rs3842689 3 F	TACTGATGCTCTCTGGTCCTGCAC	24	60.3	54.2	PrimerQuest
				PXR new rs3842689 R	CTGTAACAGGGAGCTGGTTGTATG	24	57.3	50	PrimerQuest
Intron 1	rs2472677	PCR-RFLP	621	PXR rs2472677 F	TGCTAGCAGTGCATAAGGGCTCAG	24	61	54.2	PrimerQuest
				PXR rs2472677 R	TCCTGACCTTAGGTGATCCATGCC	24	60.3	54.2	PrimerQuest
Exon 2	rs12721613 (*PXR*2*)	PCR-RFLP	408	PXR TSP LS4 F	CCGAGTTCACAGGCCCAAATGTGAGT	26	62.9	53.8	PrimerQuest
				PXR TSP LS4 R	TCAAGCTGAGGCCCTGAGACGTTAC	25	62.2	56	PrimerQuest
Intron 2	rs1464603	PCR-RFLP	522	PXR Intron 2 F	GTTCCTGGAAAGCCCAGTGTCAAC	24	60.2	54.2	PrimerQuest
				PXR Intron 2 R	TCCCTTCGGTCTGTTTCTCTTCCC	24	60.3	54.2	PrimerQuest
Intron 2	rs1464602	PCR-RFLP	522	PXR Intron 2 F	GTTCCTGGAAAGCCCAGTGTCAAC	24	60.2	54.2	PrimerQuest
				PXR Intron 2 R	TCCCTTCGGTCTGTTTCTCTTCCC	24	60.3	54.2	PrimerQuest
Intron 8	rs2276707	PCR-RFLP	291	PXR rs2276707 F	ACAGTGTTCAATGCGGAGACTGGA	24	60.4	50	PrimerQuest
				PXR rs2276707 R	TCCAGTAGAAGTTGCTGGAAGCCA	24	60.1	50	PrimerQuest

Bold lowercase letters are incorporated mutagenic nucleotides. Underlined letters are non-complementary 5'-tails.

Appendix C: List of reagents and buffers

10 mg/ml Cresol stock solution

0.1 g Cresol red (Sigma-Aldrich Co. LLC, St. Louis, USA)
1000 µl Sabax water (Adcock Ingram Pty. Ltd., Midrand, ZA)

Cresol loading dye

200 µl Cresol stock solution (10 mg/ml)
3.4 g sucrose (Merck KGaA, Darmstadt, Germany)
9.5 ml Sabax water (Adcock Ingram Pty. Ltd., Midrand, ZA)

TBE 10x buffer

108 g Tris (Sigma-Aldrich Co. LLC, St. Louis, USA)
55 g boric acid (Sigma-Aldrich Co. LLC, St. Louis, USA)
12.445 g EDTA (Sigma-Aldrich Co. LLC, St. Louis, USA)
Fill to 500 ml with distilled water (dH_2O) or water subjected to reverse osmosis (RO-water)

TBE 1x buffer

200 ml TBE 10x buffer
1.8 L distilled or RO-water

TBE 5x buffer

200 ml TBE 10x buffer
200 ml distilled or RO-water

10% (w/v) SDS solution

10 g SDS (Merck KGaA, Darmstadt, Germany)
100 ml distilled or RO-water

0.6% (w/v) Agarose gel

0.3 g agarose (Lonza Group Ltd., Rockland, USA)
50 ml distilled or RO-water

1.0% (w/v) Agarose gel

0.5 g agarose (Lonza Group Ltd., Rockland, USA)
50 ml distilled or RO-water

1.5% Agarose gel

0.75 g agarose (Lonza Group Ltd., Rockland, USA)
50 ml distilled or RO-water

3% Agarose gel

1.5 g agarose (Lonza Group Ltd., Rockland, USA)
50 ml distilled or RO-water

40% (w/v) Stock Polyacrylamide solution

38 g acrylamide (Sigma-Aldrich Co. LLC, St. Louis, USA)
2 g bisacrylamide (Promega Corporation, Fitchburg, USA)
100 ml distilled or RO-water

15% Polyacrylamide gel mix

37.5 ml stock polyacrylamide solution
20 ml 5x TBE buffer
42.5 ml distilled or RO-water

10% Ammonium persulphate solution (APS)

0.1 g ammonium persulphate (Promega Corporation, Fitchburg, USA)
100 µl distilled or RO-water

Polyacrylamide gel mix

10 ml 15% polyacrylamide gel mix
100 µl APS
20 µl TEMED (Sigma-Aldrich Co. LLC, St. Louis, USA)

Bluejuice Loading dye (adapted from Barril & Nates, 2012)

4 ml (0.25%) bromophenol blue (Merck KGaA, Darmstadt, Germany)
4 ml (0.25%) xylene cyanol FF (Sigma-Aldrich Co. LLC, St. Louis, USA)
30 ml (30%) glycerol (Merck KGaA, Darmstadt, Germany)
62 ml TBE 1x buffer

Appendix D: List of patient characteristics

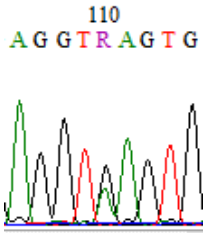
	Xhosa	CMA
Characteristic	**Median (IQR)**	**Median (IQR)**
Age at sample collection (years)	39 (34 - 45)	41 (37 - 46)
BMI Baseline	23.38 (34 - 45)	21.46 (37 - 46)
CD4 Baseline (copies / ml)	159 (98 - 206)	147 (98 - 192)
Time on EFV (months)	40.96 (24 - 57)	25.59 (10 - 41)
Characteristic	**Number of pts (%)**	**Number of pts (%)**
WHO Disease Stage:		
Stage 1	3 (3.80)	3 (5.66)
Stage 2	16 (20.25)	12 (22.64)
Stage 3	39 (49.37)	21 (39.62)
Stage 4	18 (22.78)	14 (26.42)
n/a	3 (3.80)	3 (5.66)
Smoking:		
yes	2 (2.53)	10 (18.87)
no	47 (59.49)	24 (45.28)
n/a	30 (37.97)	19 (35.85)
Alcohol:		
yes	5 (6.33)	9 (16.98)
no	48 (60.76)	26 (49.06)
n/a	26 (32.91)	18 (33.96)
Adherence		
good	70 (87.5)	43 (81.13)
acceptable	6 (7.5)	7 (13.21)
bad	3 (3.75)	2 (3.77)
n/a	1 (1.25)	1 (1.89)
Default	5 (6.33)	6 (11.32)
Change in regimen		
No change	42 (53.16)	27 (50.94)
Change	37 (46.84)	26 (49.06)
More than once	9 (11.39)	7 (13.21)
Regimen during sampling		
EFV/3TC/d4T	46 (58.23)	33 (62.26)
EFV/3TC/AZT	21 (26.58)	9 (16.98)
EFV/3TC/TDF	12 (15.19)	10 (18.87)
n/a	0 (0.00)	1 (1.89)
Second-line therapy		
yes	6 (7.59)	1 (1.89)
no	73 (92.41)	52 (98.11)
n/a	0 (0.00)	1 (1.89)
Additional medication		
yes	18 (22.78)	13 (24 53)
no	61 (77.22)	40 (75.47)

Appendix E: Regions of STRs in Amplicons UA2 and UA3

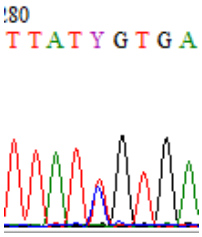
The red coloured text regions indicate sites of STRs, resulting in enzyme slippage in a sequencing reaction. Blue shaded regions indicate the UA2 primer set. Green shaded regions indicate the UA3 primer set. The region shaded grey indicates the forward primer for Exon 1.

Appendix F: Chromatograms of SNPs identified through bi-directional sequencing.

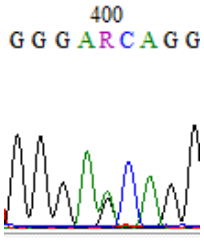
rs55802895 (258G>A)



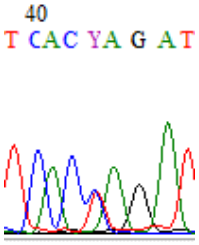
rs2501870 (432C>T)



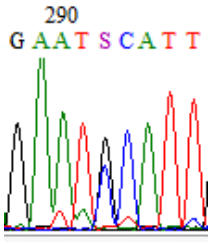
rs11584174 (548G>A)



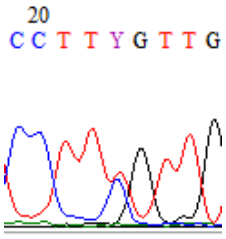
rs35709078 (7145C>T)



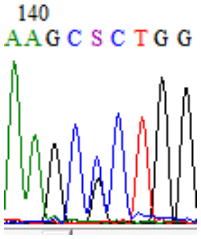
rs141260451 (7393C>G)



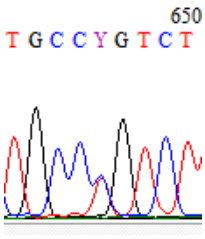
rs2502815 (9774C>T)



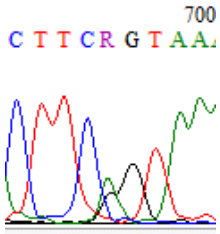
rs35205211 (9892C>G)



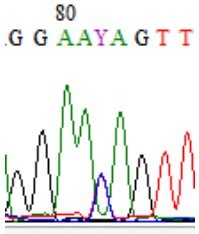
rs2307424 (10396C>T)



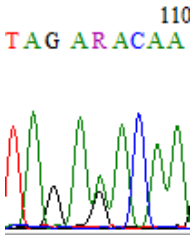
rs2307420 (10445A>G)



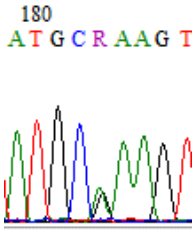
rs34859188 (13196T>C)



rs35103872 (13220G>A)



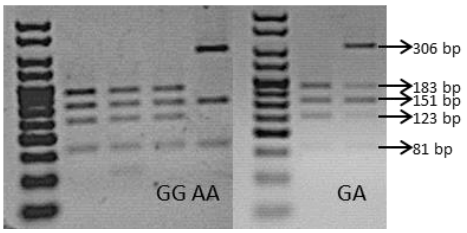
rs36017137 (13297G>A)



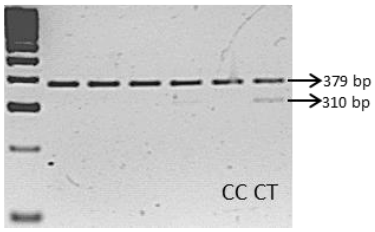
Appendix G: Gel photos of SNP genotypes

NR1I3 SNPs:

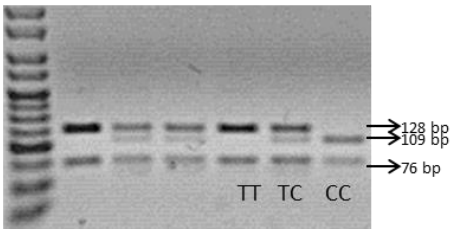
rs55802895 (258G>A)



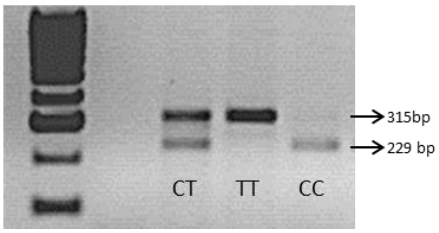
rs35709078 (7145C>T)



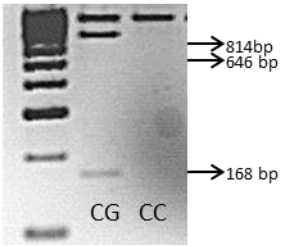
rs3003596 (8784T>C)



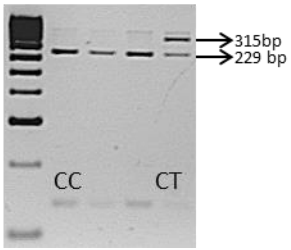
rs2502815 (9774C>T)



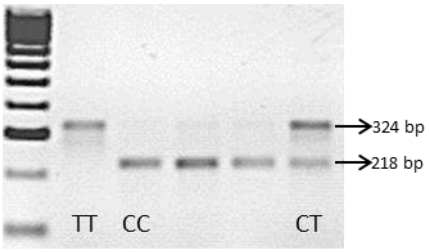
rs35205211 (9892C>G)



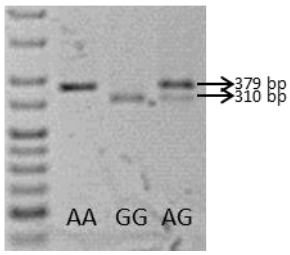
rs34161743 (9923C>T)



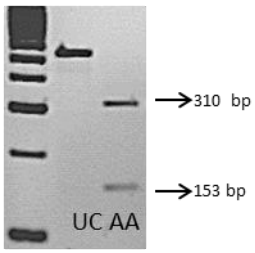
rs2307424 (10396C>T)



rs2307420 (10445A>G)

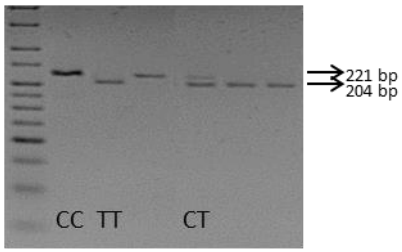


rs113800307 (12076A>G)

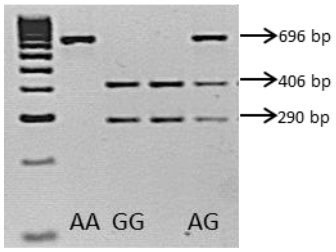


NR1I2 SNPs:

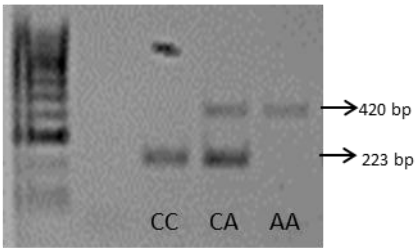
rs3814055 (5705C>T)



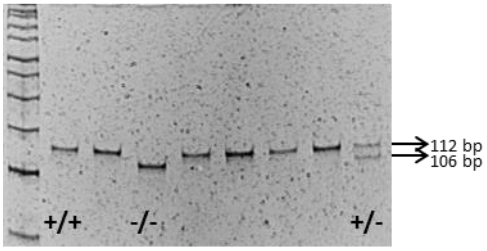
rs1523128 (6334G>A)



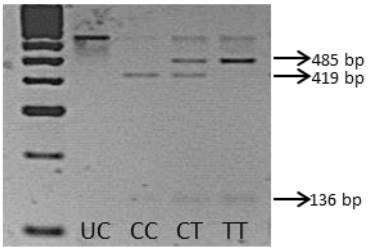
rs1523127 (6709C>A)



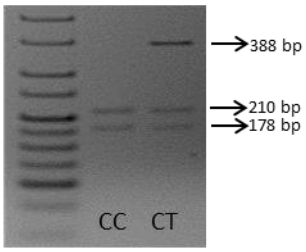
rs3842689
(7070_7075delGAGAAG)



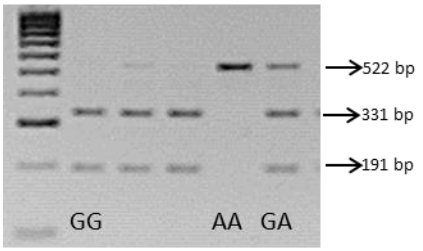
rs2472677 (24087C>T)



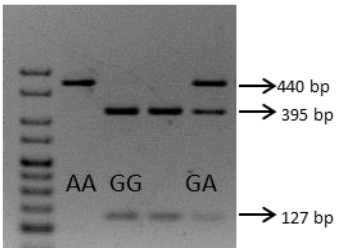
rs12721613 (31846C>T)



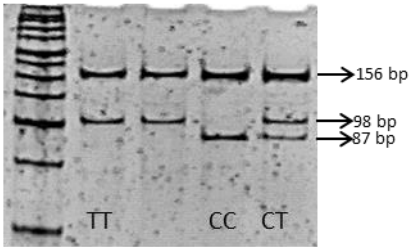
rs1464603 (32019G>A)



rs1464602 (32042G>A)



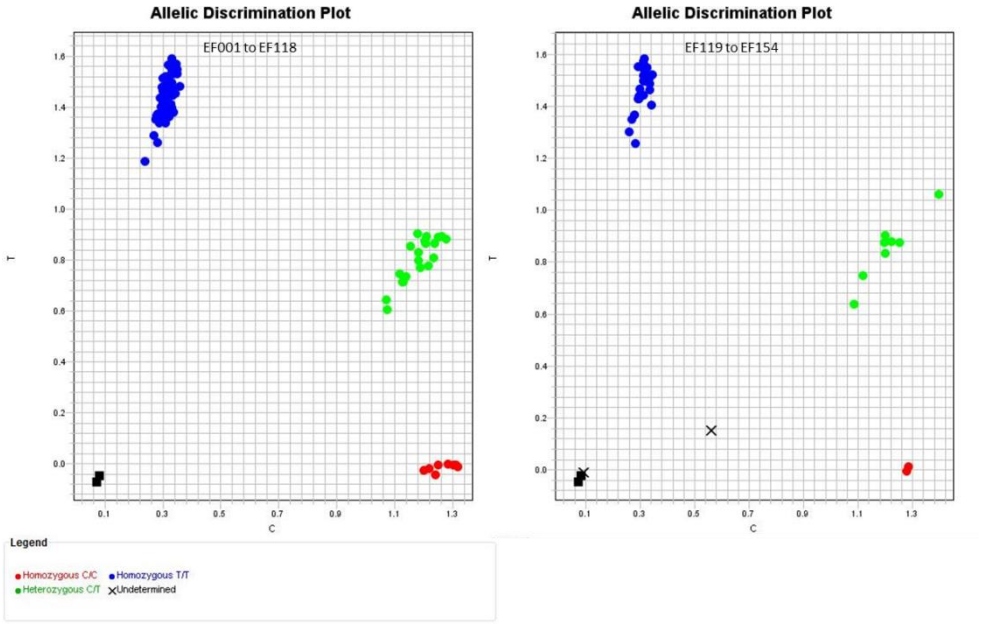
rs2276707 (39823C>T)



Appendix H: Taqman® analysis plots for rs1523130 (5177T>C)

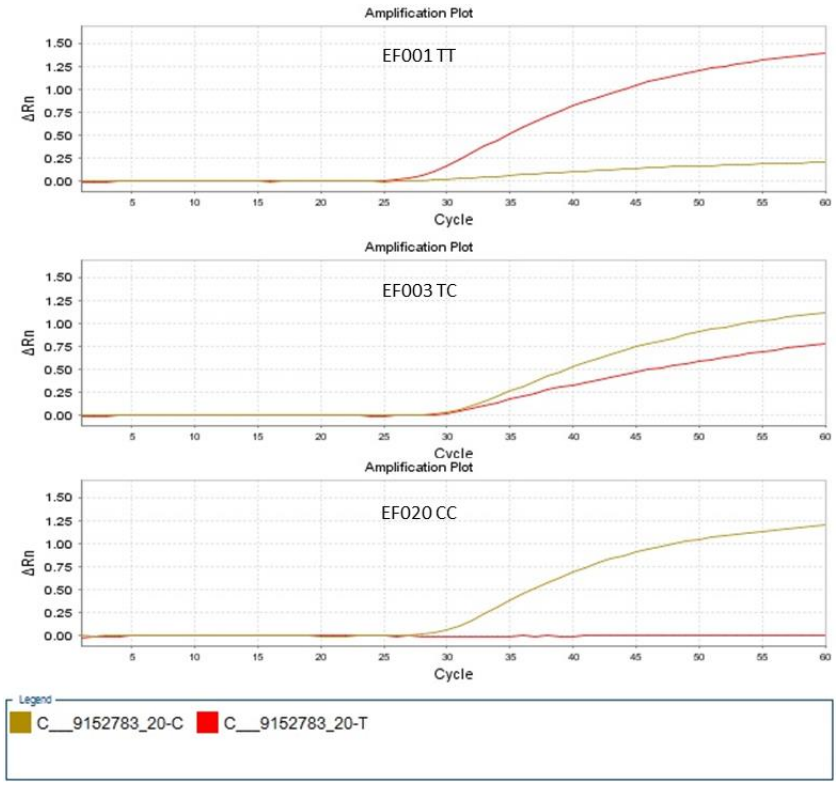
Allelic discrimination plots:

Allelic discrimination plots for all samples showing the change in fluorescence for the T-allele probe on the y-axis and the C-allele probe on the x-axis.



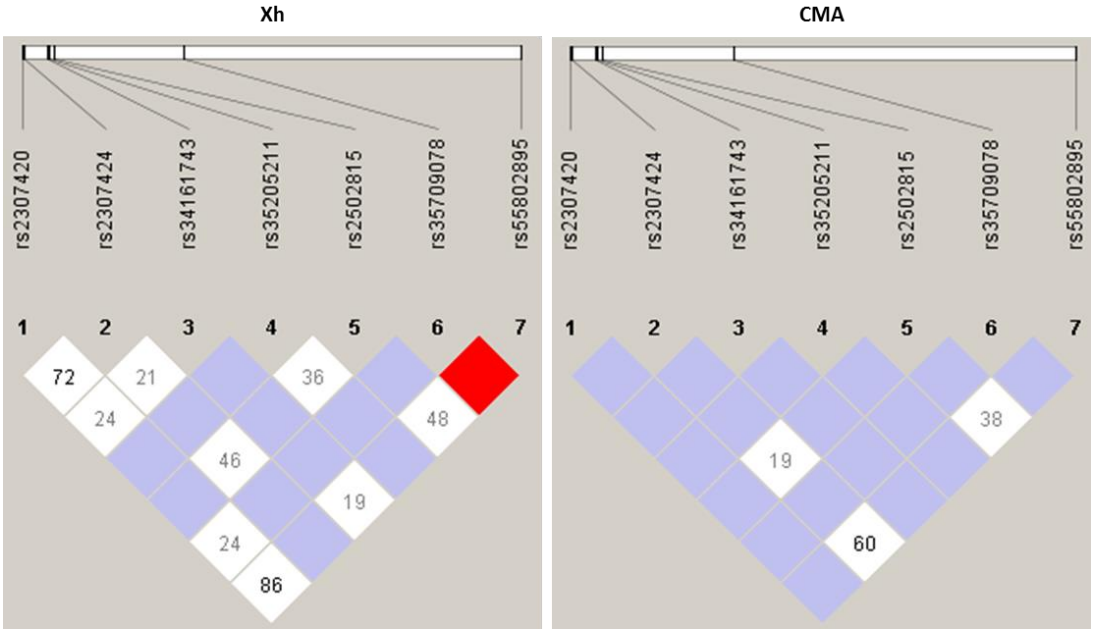
Amplification plots:

Amplification plots for three samples showing the change in fluorescence per PCR cycle for the two fluorescent probes specific for each allele, normalised to baseline.

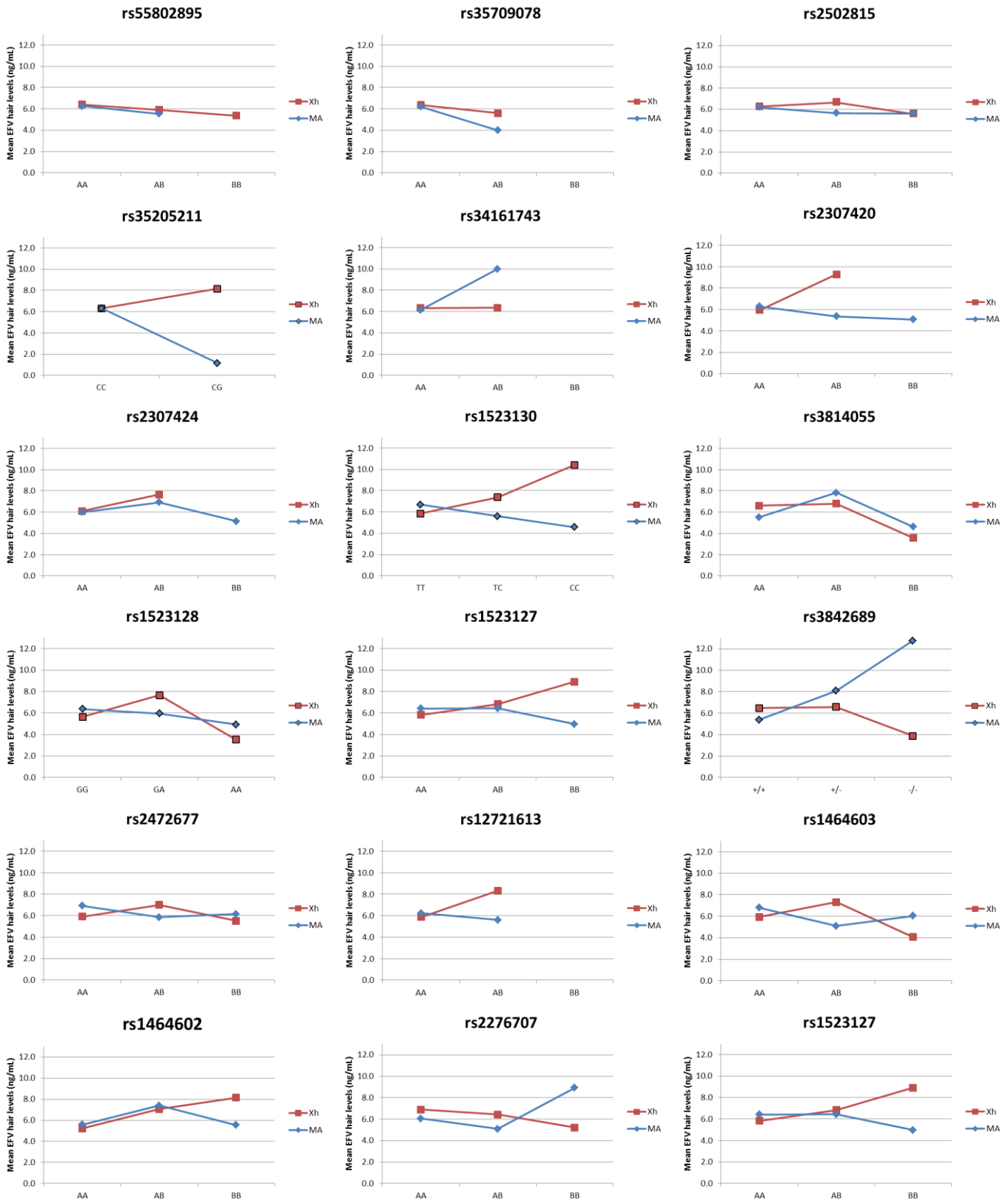


Appendix I: LD plots for *NR1I3* variants in the Xhosa and CMA cohorts

LD values are represented in the blocks by correlation coefficient (r2) values.



Appendix J: Mean EFV levels for all *NR1I3* and *NR1I2* variants



Appendix K: Tables of allele frequencies according to treatment outcome and the presence of adverse drug reactions.

Table K-1: Genotype frequencies and MAFs for *NR1I3* and *NR1I2* variants in the Xhosa and CMA cohorts for responders and non-responders

Gene	rs number	Genomic identification	Genotype	Xhosa				CMA			
				Responders		Non-Responders		Responders		Non-Responders	
				n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF
NR1I3	rs2307420	10445A>G	AA	64 (0.914)	0.028	3 (0.600)	0.083	31 (0.861)	0.028	6 (0.857)	0.042
			AG	6 (0.086)		2 (0.040)		4 (0.111)		1 (0.143)	
			GG	0 (0.000)		0 (0.000)		1 (0.028)		0 (0.000)	
NR1I3	rs2307424	10396C>T	CC	60 (0.857)	0.047	4 (0.800)	0.042	27 (0.750)	0.052	4 (0.571)	0.167
			CT	10 (0.143)		1 (0.200)		7 (0.194)		2 (0.286)	
			TT	0 (0.000)		0 (0.000)		2 (0.056)		1 (0.143)	
NR1I3	rs34161743	9923C>T	CC	67 (0.957)	0.014	5 (1.000)	0.000	35 (1.000)	0.000	6 (0.857)	0.042
			CT	3 (0.043)		0 (0.000)		0 (0.000)		1 (0.143)	
			TT	0 (0.000)		0 (0.000)		0 (0.000)		0 (0.000)	
NR1I3	rs35205211	9892C>G	CC	66 (0.985)	0.005	5 (1.000)	0.000	35 (0.972)	0.005	7 (1.000)	0.000
			CG	1 (0.015)		0 (0.000)		1 (0.028)		0 (0.000)	
			GG	0 (0.000)		0 (0.000)		0 (0.000)		0 (0.000)	
NR1I3	rs2502815	9774C>T	CC	41 (0.594)	0.156	1 (0.200)	0.208	17 (0.486)	0.104	3 (0.429)	0.208
			CT	23 (0.333)		3 (0.600)		14 (0.400)		3 (0.429)	
			TT	5 (0.072)		1 (0.200)		4 (0.114)		1 (0.143)	
NR1I3	rs35709078	7145C>T	CC	68 (0.971)	0.009	4 (0.800)	0.042	35 (0.972)	0.005	7 (1.000)	0.000
			CT	2 (0.029)		1 (0.200)		1 (0.028)		0 (0.000)	
			TT	0 (0.000)		0 (0.000)		0 (0.000)		0 (0.000)	
NR1I3	rs55802895	258G>A	GG	59 (0.843)	0.057	4 (0.800)	0.042	33 (0.917)	0.014	5 (0.714)	0.083
			GA	10 (0.143)		1 (0.200)		3 (0.083)		2 (0.286)	
			AA	1 (0.014)		0 (0.000)		0 (0.000)		0 (0.000)	
NR1I2	rs1523130	5177T>C	TT	54 (0.783)	0.087	5 (1.000)	0.000	20 (0.571)	0.096	5 (0.714)	0.125
			TC	12 (0.174)		0 (0.000)		10 (0.286)		1 (0.143)	

Gene	rs number	Genomic identification	Genotype	Xhosa				CMA			
				Responders		Non-Responders		Responders		Non-Responders	
				n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF
			CC	3 (0.043)		0 (0.000)		5 (0.143)		1 (0.143)	
NR1I2	**rs3814055**	**5705C>T**	**CC**	**47 (0.701)**		**0 (0.000)**		27 (0.771)		5 (0.714)	
			CT	**16 (0.239)**	**0.113**	**3 (0.750)**	**0.208**	7 (0.200)	0.042	2 (0.286)	0.083
			TT	**4 (0.060)**		**1 (0.250)**		1 (0.029)		0 (0.000)	
NR1I2	rs1523128	6334G>A	GG	34 (0.493)		4 (0.800)		24 (0.667)		3 (0.429)	
			GA	30 (0.435)	0.190	1 (0.200)	0.042	11 (0.306)	0.062	3 (0.429)	0.208
			AA	5 (0.072)		0 (0.000)		1 (0.028)		1 (0.143)	
NR1I2	rs1523127	6709C>A	CC	46 (0.676)		2 (0.400)		15 (0.429)		3 (0.429)	
			CA	19 (0.279)	0.120	2 (0.400)	0.167	14 (0.400)	0.125	3 (0.429)	0.208
			AA	3 (0.044)		1 (0.200)		6 (0.171)		1 (0.143)	
NR1I2	rs3842689	7070_7075delGAGAAG	+/+	47 (0.681)		1 (0.429)		26 (0.722)		5 (0.714)	
			+/-	18 (0.261)	0.123	3 (0.429)	0.208	8 (0.222)	0.057	2 (0.286)	0.083
			-/-	4 (0.058)		1 (0.143)		2 (0.056)		0 (0.000)	
NR1I2	**rs2472677**	**24087C>T**	**CC**	27 (0.391)		2 (0.400)		**18 (0.514)**		**0 (0.000)**	
			CT	30 (0.435)	0.260	3 (0.600)	0.125	**13 (0.371)**	**0.101**	**6 (0.857)**	**0.333**
			TT	12 (0.174)		0 (0.000)		**4 (0.114)**		**1 (0.143**	
NR1I2	rs12721613	31846C>T	CC	55 (0.797)		4 (0.800)		33 (0.917)		7 (1.000)	
			CT	14 (0.203)	0.066	1 (0.200)	0.042	3 (0.083)	0.014	0 (0.000)	0.000
			TT	0 (0.000)		0 (0.000)		0 (0.000)		0 (0.000)	
NR1I2	rs1464603	32019G>A	GG	45 (0.643)		3 (0.600)		20 (0.556)		4 (0.571)	
			GA	24 (0.343)	0.123	1 (0.200)	0.125	14 (0.389)	0.085	2 (0.286)	0.167
			AA	1 (0.014)		1 (0.200)		2 (0.056)		1 (0.143)	
NR1I2	rs1464602	32042G>A	GG	31 (0.443)		2 (0.400)		12 (0.353)		4 (0.571)	
			GA	31 (0.443)	0.226	2 (0.400)	0.167	13 (0.382)	0.149	2 (0.286)	0.167
			AA	8 (0.114)		1 (0.200)		9 (0.265)		1 (0.143)	

Gene	rs number	Genomic identification	Genotype	Xhosa				CMA			
				Responders		Non-Responders		Responders		Non-Responders	
				n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF
NR1I2	rs2276707	39823C>T	CC	21 (0.304)	0.295	1 (0.200)	0.208	18 (0.514)	0.110	2 (0.286)	0.250
			CT	34 (0.493)		3 (0.600)		11 (0.314)		4 (0.571)	
			TT	14 (0.203)		1 (0.200)		6 (0.171)		1 (0.143)	

Genotype frequencies that differ significantly between responders and non-responders are indicated in bold.

Table K-2: Genotype frequencies and MAFs for *NR1I3* and *NR1I2* variants in the Xhosa and CMA cohorts for ADR and Non-ADR groups

rs number	Genomic identification	Genotype	Xhosa				CMA			
			Non-ADR		ADR		Non-ADR		ADR	
			n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF
rs2307420	10445A>G	AA	61 (0.897)	0.051	9 (0.818)	0.091	43 (0.896)	0.063	4 (0.800)	0.100
		AG	7 (0.103)		2 (0.182)		4 (0.083)		1 (0.200)	
		GG	0 (0.000)		0 (0.000)		1 (0.021)		0 (0.000)	
rs2307424	10396C>T	CC	58 (0.853)	0.074	9 (0.818)	0.091	35 (0.729)	0.156	2 (0.400)	0.400
		CT	10 (0.147)		2 (0.182)		11 (0.229)		2 (0.400)	
		TT	0 (0.000)		0 (0.000)		2 (0.042)		1 (0.200)	
rs34161743	9923C>T	CC	65 (0.956)	0.022	11 (1.000)	0.000	46 (0.979)	0.011	5 (1.000)	0.000
		CT	3 (0.044)		0 (0.000)		1 (0.021)		0 (0.000)	
		TT	0 (0.000)		0 (0.000)		0 (0.000)		0 (0.000)	
rs35205211	9892C>G	CC	64 (0.985)	0.008	10 (0.909)	0.045	46 (0.979)	0.011	5 (1.000)	0.000
		CG	1 (0.015)		1 (0.091)		1 (0.021)		0 (0.000)	
		GG	0 (0.000)		0 (0.000)		0 (0.000)		0 (0.000)	
rs2502815	9774C>T	CC	39 (0.574)	0.257	4 (0.400)	0.350	22 (0.468)	0.319	2 (0.400)	0.300
		CT	23 (0.338)		5 (0.500)		20 (0.426)		3 (0.600)	
		TT	6 (0.088)		1 (0.100)		5 (0.106)		0 (0.000)	
rs35709078	7145C>T	CC	65 (0.956)	0.022	11 (1.000)	0.000	47 (0.979)	0.010	5 (1.000)	0.000
		CT	3 (0.044)		0 (0.000)		1 (0.021)		0 (0.000)	
		TT	0 (0.000)		0 (0.000)		0 (0.000)		0 (0.000)	
rs55802895	258G>A	GG	57 (0.838)	0.088	10 (0.909)	0.045	42 (0.875)	0.063	4 (0.800)	0.100
		GA	10 (0.147)		1 (0.091		6 (0.125)		1 (0.200)	
		AA	1 (0.015)		0 (0.000)		0 (0.000)		0 (0.000)	

rs number	Genomic identification	Genotype	Xhosa				CMA			
			Non-ADR		ADR		Non-ADR		ADR	
			n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF
rs1523130	5177T>C	TT	53 (0.791)		8 (0.727)		28 (0.583)		3 (0.750)	
		TC	12 (0.179)	0.119	2 (0.182)	0.182	13 (0.271)	0.281	1 (0.250)	0.125
		CC	2 (0.030)		1 (0.091)		7 (0.146)		0 (0.000)	
rs3814055	5705C>T	CC	44 (0.677)		4 (0.404)		37 (0.787)		5 (1.000)	
		CT	16 (0.246)	0.200	5 (0.500)	0.350	9 (0.191)	0.117	0 (0.000)	0.000
		TT	5 (0.077)		1 (0.100)		1 (0.021)		0 (0.000)	
rs1523128	6334G>A	AA	33 (0.493)		8 (0.727)		31 (0.646)		2 (0.400)	
		AG	29 (0.433)	0.291	3 (0.273)	0.136	15 (0.313)	0.198	3 (0.600)	0.300
		GG	5 (0.075)		0 (0.000)		2 (0.042)		0 (0.000)	
rs1523127	6709C>A	CC	43 (0.652)		7 (0.198)		20 (0.426)		3 (0.600)	
		CA	20 (0.303)	0.197	3 (0.273)	0.227	18 (0.383)	0.383	2 (0.400)	0.200
		AA	3 (0.045)		1 (0.091)		9 (0.191)		0 (0.000)	
rs3842689	7070_7075del GAGAAG	+/+	45 (0.672)		5 (0.455)		37 (0.771)		4 (0.800)	
		+/-	18 (0.269)	0.194	5 (0.455)	0.318	9 (0.188)	0.135	1 (0.200)	0.100
		-/-	4 (0.060)		1 (0.091)		2 (0.042)		0 (0.000)	
rs2472677	24087C>T	CC	28 (0.418)		2 (0.182)		15 (0.333)		3 (0.600)	
		CT	30 (0.448)	0.358	5 (0.455)	0.591	25 (0.556)	0.389	2 (0.400)	0.200
		TT	9 (0.134)		4 (0.364)		5 (0.111)		0 (0.000)	
rs12721613	31846C>T	CC	54 (0.806)		9 (0.818)		44 (0.917)		5 (1.000)	
		CT	13 (0.194)	0.097	2 (0.182)	0.091	4 (0.083)	0.042	0 (0.000)	0.000
		TT	0 (0.000)		0 (0.000)		0 (0.000)		0 (0.000)	
rs1464603	32019G>A	GG	46 (0.677)		5 (0.455)		29 (0.604)		3 (0.600)	
		GA	20 (0.294)	0.176	6 (0.545)	0.273	16 (0.333)	0.229	2 (0.400)	0.200

rs number	Genomic identification	Genotype	Xhosa				CMA			
			Non-ADR		ADR		Non-ADR		ADR	
			n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF	n (frequency)	MAF
		AA	2 (0.029)		0 (0.000)		3 (0.063)		0 (0.000)	
		GG	31 (0.456)		5 (0.455)		19 (0.413)		2 (0.400)	
rs1464602	32042G>A	GA	28 (0.412)	0.273	6 (0.545)	0.273	17 (0.370)	0.402	2 (0.400)	0.400
		AA	9 (0.132)		0 (0.000)		10 (0.217)		1 (0.200)	
		CC	17 (0.254)		6 (0.545)		21 (0.447)		2 (0.400)	
rs2276707	39823C>T	CT	37 (0.552)	0.470	3 (0.273)	0.318	17 (0.632)	0.372	3 (0.600)	0.300
		TT	13 (0.194)		2 (0.182)		9 (0.191)		0 (0.000)	

Appendix L: Association analysis between composite genotypes and efavirenz levels using logistic regression models

Gene	rs number	Genotypic identification	*P*-value	Quantitative effect on EFV levels
***CYP2B6*6 (516G>T)**				
NR1I3	rs2307420	10445A>G	0.393	n/a
NR1I3	rs2307424	10396C>T	0.856	n/a
NR1I3	rs34161743	9923C>T	n/a	n/a
NR1I3	rs35205211	9892C>G	n/a	n/a
NR1I3	rs2502815	9774C>T	0.945	n/a
NR1I3	rs3003596	8784T>C	0.192	n/a
NR1I3	rs35709078	7145C>T	0.617	n/a
NR1I3	rs55802895	258G>A	0.005	161% ↑ from *6*6|258GG to Non-*6*6|258GG No significant difference from *6*6|258GA to Non-*6*6|258GA
NR1I2	rs1523130	5177T>C	0.778	n/a
NR1I2	rs3814055	5705C>T	0.783	n/a
NR1I2	rs1523128	6334G>A	0.524	n/a
NR1I2	rs1523127	6709C>A	0.401	n/a
NR1I2	rs3842689	7070_7075delG AGAAG	0.755	n/a
NR1I2	rs2472677	24087C>T	0.634	n/a
NR1I2	rs12721613	31846C>T	0.365	n/a
NR1I2	rs1464603	32019G>A	0.331	n/a
NR1I2	rs1464602	32042G>A	0.361	n/a
NR1I2	rs2276707	39823C>T	0.877	n/a
***CYP1A2*1F* (163C>A)**				
NR1I3	rs2307420	10445A>G	0.681	n/a
NR1I3	rs2307424	10396C>T	0.152	n/a
NR1I3	rs34161743	9923C>T	0.648	n/a
NR1I3	rs35205211	9892C>G	0.068	n/a
NR1I3	rs2502815	9774C>T	0.026	436% ↑ from 163CC|9774CT to 163CC|977CC 390% ↑ from 163CC|9774TT to 163CC|977CC
NR1I3	rs3003596	8784T>C	0.316	n/a
NR1I3	rs35709078	7145C>T	n/a	n/a
NR1I3	rs55802895	258G>A	0.269	n/a
NR1I2	rs1523130	5177T>C	0.130	n/a
NR1I2	rs3814055	5705C>T	0.122	n/a
NR1I2	rs1523128	6334G>A	0.635	n/a
NR1I2	rs1523127	6709C>A	0.153	n/a
NR1I2	rs3842689	7070_7075delG AGAAG	0.017	36% ↑ from 163AC|+/- to 163AC|+/+ 263% ↑ from 163|-/- to 163AC|+/+
NR1I2	rs2472677	24087C>T	0.354	n/a
NR1I2	rs12721613	31846C>T	0.026	86% ↓ from 163CC|31846CT to 163AA|31846CT
NR1I2	rs1464603	32019G>A	0.682	n/a
NR1I2	rs1464602	32042G>A	0.210	n/a
NR1I2	rs2276707	39823C>T	0.030	no definable pattern detected
***CYP2A6*9B* (48T>G)**				
NR1I3	rs2307420	10445A>G	0.007	464% ↑ from 48 TG|10445AA to 48 TG|10445AG

NR1I3	rs2307424	10396C>T	n/a	n/a
NR1I3	rs34161743	9923C>T	n/a	n/a
NR1I3	rs35205211	9892C>G	n/a	n/a
NR1I3	rs2502815	9774C>T	0.470	n/a
NR1I3	rs3003596	8784T>C	0.599	n/a
NR1I3	rs35709078	7145C>T	n/a	n/a
NR1I3	rs55802895	258G>A	n/a	n/a
NR1I2	rs1523130	5177T>C	0.483	n/a
NR1I2	rs3814055	5705C>T	0.605	n/a
NR1I2	rs1523128	6334G>A	0.984	n/a
NR1I2	rs1523127	6709C>A	0.121	n/a
NR1I2	rs3842689	7070_7075delG AGAAG	0.409	n/a
NR1I2	rs2472677	24087C>T	0.872	n/a
NR1I2	rs12721613	31846C>T	n/a	n/a
NR1I2	rs1464603	32019G>A	0.021	215% ↑ from 48TG|32019GG to 48TG|32019AG
NR1I2	rs1464602	32042G>A	0.046	465% ↑ from 48TG|32042GG to 48TG|32042GA
NR1I2	rs2276707	39823C>T	0.676	n/a
***CYP3A5*6* (14690G>A)**				
NR1I3	rs2307420	10445A>G	0.207	n/a
NR1I3	rs2307424	10396C>T	0.960	n/a
NR1I3	rs34161743	9923C>T	0.650	n/a
NR1I3	rs35205211	9892C>G	n/a	n/a
NR1I3	rs2502815	9774C>T	0.911	n/a
NR1I3	rs3003596	8784T>C	0.907	n/a
NR1I3	rs35709078	7145C>T	n/a	n/a
NR1I3	rs55802895	258G>A	0.590	n/a
NR1I2	rs1523130	5177T>C	0.247	n/a
NR1I2	rs3814055	5705C>T	0.639	n/a
NR1I2	rs1523128	6334G>A	0.312	n/a
NR1I2	rs1523127	6709C>A	0.118	n/a
NR1I2	rs3842689	7070_7075delG AGAAG	0.426	n/a
NR1I2	rs2472677	24087C>T	0.289	n/a
NR1I2	rs12721613	31846C>T	0.353	n/a
NR1I2	rs1464603	32019G>A	0.888	n/a
NR1I2	rs1464602	32042G>A	0.907	n/a
NR1I2	rs2276707	39823C>T	0.825	n/a
***CYP2A6*17* (5065G>A)**				
NR1I3	rs2307420	10445A>G	0.172	n/a
NR1I3	rs2307424	10396C>T	0.688	n/a
NR1I3	rs34161743	9923C>T	n/a	n/a
NR1I3	rs35205211	9892C>G	n/a	n/a
NR1I3	rs2502815	9774C>T	0.442	n/a
NR1I3	rs3003596	8784T>C	0.700	n/a
NR1I3	rs35709078	7145C>T	0.424	n/a
NR1I3	rs55802895	258G>A	0.266	n/a
NR1I2	rs1523130	5177T>C	0.971	n/a
NR1I2	rs3814055	5705C>T	0.832	n/a
NR1I2	rs1523128	6334G>A	0.904	n/a

NR1I2	rs1523127	6709C>A	0.990	n/a
NR1I2	rs3842689	7070_7075delGAGAAG	0.723	n/a
NR1I2	rs2472677	24087C>T	0.516	n/a
NR1I2	rs12721613	31846C>T	0.090	n/a
NR1I2	rs1464603	32019G>A	0.743	n/a
NR1I2	rs1464602	32042G>A	0.003	no definable pattern detected
NR1I2	rs2276707	39823C>T	0.813	n/a
***CYP3A4*1B* (392A>G)**				
NR1I3	rs2307420	10445A>G	0.664	n/a
NR1I3	rs2307424	10396C>T	0.975	n/a
NR1I3	rs34161743	9923C>T	0.673	n/a
NR1I3	rs35205211	9892C>G	0.809	n/a
NR1I3	rs2502815	9774C>T	0.587	n/a
NR1I3	rs3003596	8784T>C	0.569	n/a
NR1I3	rs35709078	7145C>T	0.602	n/a
NR1I3	rs55802895	258G>A	0.665	n/a
NR1I2	rs1523130	5177T>C	0.011	64% ↑ from 392GG|5177TT to 392GG|5177TC 66% ↓ from 392AG|5177TT to 392AG|5177CC
NR1I2	rs3814055	5705C>T	0.246	n/a
NR1I2	rs1523128	6334G>A	0.686	n/a
NR1I2	rs1523127	6709C>A	0.093	n/a
NR1I2	rs3842689	7070_7075delGAGAAG	0.084	n/a
NR1I2	rs2472677	24087C>T	0.658	n/a
NR1I2	rs12721613	31846C>T	0.686	n/a
NR1I2	rs1464603	32019G>A	0.961	n/a
NR1I2	rs1464602	32042G>A	0.963	n/a
NR1I2	rs2276707	39823C>T	0.125	n/a
***CYP3A5*3* (6986A>G)**				
NR1I3	rs2307420	10445A>G	0.077	n/a
NR1I3	rs2307424	10396C>T	0.495	n/a
NR1I3	rs34161743	9923C>T	0.641	n/a
NR1I3	rs35205211	9892C>G	0.060	n/a
NR1I3	rs2502815	9774C>T	0.572	n/a
NR1I3	rs3003596	8784T>C	0.436	n/a
NR1I3	rs35709078	7145C>T	n/a	n/a
NR1I3	rs55802895	258G>A	0.461	n/a
NR1I2	rs1523130	5177T>C	0.669	n/a
NR1I2	rs3814055	5705C>T	0.056	n/a
NR1I2	rs1523128	6334G>A	0.265	n/a
NR1I2	rs1523127	6709C>A	0.649	n/a
NR1I2	rs3842689	7070_7075delGAGAAG	0.008	105% ↑ from 6986AG|+/+ to 6986AG|+/- 199% ↑ from 6986AG|+/+ to 6986AG|-/-
NR1I2	rs2472677	24087C>T	0.834	n/a
NR1I2	rs12721613	31846C>T	n/a	n/a
NR1I2	rs1464603	32019G>A	0.068	n/a
NR1I2	rs1464602	32042G>A	0.229	n/a
NR1I2	rs2276707	39823C>T	0.023	126% ↑ from 6986AG|39823CC to 6986AG|39823TT

Appendix M: Analysis of the effect of genotype on treatment outcome and adverse reactions

Gene	rs number	Genomic identification	*P*-Value	
			Treatment Outcome	ADRs
NR1I3	rs2307420	10445A>G	0.2893	0.4724
NR1I3	rs2307424	10396C>T	0.4911	0.2902
NR1I3	rs34161743	9923C>T	0.5136	0.3292
NR1I3	rs35205211	9892C>G	0.6674	0.3244
NR1I3	rs2502815	9774C>T	0.1831	0.2244
NR1I3	rs3003596	8784T>C	0.7856	0.2845
NR1I3	rs35709078	7145C>T	0.0631	0.5152
NR1I3	rs55802895	258G>A	0.3011	0.9351
NR1I2	rs1523130	5177T>C	0.1842	0.9486
NR1I2	rs3814055	5705C>T	0.3843	0.6089
NR1I2	rs1523128	6334G>A	0.9342	0.0675
NR1I2	rs1523127	6709C>A	0.8016	0.9903
NR1I2	rs3842689	7070_7075delGAGAAG	0.4283	0.4756
NR1I2	rs2472677	24087C>T	0.1180	0.3833
NR1I2	rs12721613	31846C>T	0.9555	0.4012
NR1I2	rs1464603	32019G>A	0.7220	0.0958
NR1I2	rs1464602	32042G>A	0.1260	0.3809
NR1I2	rs2276707	39823C>T	0.2316	0.5780