

https://dspace.org/

Genetic variants of d4T drug transporters and dNTP pool regulators, and their association with response to d4T-ART

Moketla, Blessings Marvin

2017

https://hdl.handle.net/10539/24022

Moketla, Blessings Marvin (2017) Genetic variants of d4T drug transporters and dNTP pool regulators, and their association with response to d4T-ART, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/24022>

Downloaded from a DSpace repository. https://dspace.org/





GENETIC VARIANTS OF d4T DRUG TRANSPORTERS AND dNTP POOL REGULATORS, AND THEIR ASSOCIATION WITH RESPONSE TO d4T-ART.

MALEBO BLESSINGS MARVIN MOKETLA

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Genetics.

SUPERVISED BY DR DEBRA DE ASSIS ROSA

JOHANNESBURG, SOUTH AFRICA

2017

DECLARATION

I, Marvin Blessings Malebo Moketla (554353), am a student registered for the degree of Master's in Genetics in the academic year 2015-2016.

I hereby declare the following:

- **I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.**
- **I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.**
- **I have not submitted this work before for any other degree or examination at this or any other University.**
- **The information used in the Thesis/Dissertation/Research Report HAS NOT been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.**
- **I have followed the required conventions in referencing the thoughts and ideas of others.**
- **I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.**

NRF declaration

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

Signature  **day of 24 August 2017**

DEDICATION

THIS THESIS IS DEDICATED TO MY FAMILY AND FRIENDS.

ABSTRACT

Background: Stavudine (d4T) use is associated with the development of sensory neuropathy (SN), several mechanisms may underlie d4T-induced toxicity, including:

(1) Inter-patient genetic variability in the genes modulating the deoxynucleotide triphosphate (dNTP) pool sizes.

(2) Variation in intracellular ARV drug concentrations due to genetic variation in drug transporters.

In our study we examined the genetic variation in four stavudine transporter genes and seven genes regulating the deoxythymidine triphosphate (dTTP) synthesis and their associations with d4T-induced SN or CD4+ T cell count or mtDNA copy number.

Methods: We examined a cohort of HIV-positive South African (SA) adults exposed to d4T, including 143 cases with SN and 120 controls without SN. 26 single nucleotide polymorphisms (SNPs) from the literature were chosen, prioritised on being tagSNPs with minor allele frequency >5% in Kenyan Luhya (a proxy population for the SA Black population); SNP functional effects and suitability for multiplex analysis on the genotyping platform. Genotyping was performed using Sequenom mass spectrometry. A qPCR assay was used to measure the mtDNA copy number. Association of sensory neuropathy, CD4+ T cell count and mtDNA copy number with genetic variants was evaluated using PLINK.

Results: All 26 SNPs were in Hardy-Weinberg equilibrium (HWE) in both the cases and controls. SNP rs8187758 of the *SLC28A1* transporter gene and a 3-SNP haplotype *ABCG2* were significantly associated with CD4+ T cell count after correction for multiple testing ($p = 0.043$ and $p=0.042$ respectively), but were not significant in multivariate testing. No SNP remained significantly associated with SN or mtDNA copy number, after correction for multiple testing.

Conclusion: Variation in genes encoding molecular transporters of d4T may influence CD4+ T cell counts after ART. This study presents a positive step towards achieving personalized medicine in SA.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my Lord and Saviour Jesus Christ, for the strength, grace and ability that have enabled me to endure throughout my masters.

I would like to appreciate and thank my supervisor Dr de Assis Rosa, for the opportunity she has given me to come and pursue my Masters with her, I am honoured to have been one of your masters' student in the field that I love so much. I have learned a lot through the platform that she has exposed me to and I am grateful for that. The meetings we held and the discussions we had were very productive and stimulating, I have acquired so much from those and they have tuned me to think critically. I am thankful for the amount of work and energy she dedicated to my MSc her ideas and contributions were very central to the completion of my studies.

I would also like to thank my advisor Dr Ntwasa for his input in my proposal and his willingness to assist and advice in my research are greatly appreciated.

I thank members of our lab they were instrumental in sharpening and challenging my views and for their camaraderie.

I am very thankful to my colleague Tebogo Marutha for the collaboration we had and for the help with the qPCR assay.

I would also like to thank Mr Tyrone Otgaar and Dr Boitelo Letsolo for the help with cell culturing.

I would also like to thank Sibusiso Malindisa for the help with the qPCR.

I would also like to thank Dr Nikki Gentle for the introductory training and assistance with PLINK.

I would like to thank my family for the support and encouragement, they were very helpful.

I would like to thank the National Research Foundation (NRF) for financial support.

TABLE OF CONTENTS

DECLARATION ii
DEDICATION iii
ABSTRACT iv
ACKNOWLEDGEMENTS v
TABLE OF CONTENTS vi
LIST OF FIGURES x
LIST OF TABLES xi
LIST OF ABBREVIATIONS xiii
CONFERENCE PRESENTATIONS ARISING FROM THIS STUDY xvii
CHAPTER 1: INTRODUCTION 2
1.1 HIV and HAART 2
1.2 Nucleoside reverse transcriptase inhibitor (NRTIs) 4
1.3 Stavudine 6
1.4 Mitochondrial toxicity of NRTIs 6
1.5 Sensory neuropathy 10
1.6 Pharmacogenetics of SN 11
1.7 dNTP synthesis and metabolism pathways 12
1.7.1 *de novo* synthesis of pyrimidines 13
1.7.2 Folate cycle 17
1.7.3 Salvage pathway of dNTP production 19
1.8 Transporter proteins 21
1.8.1 Transport of nucleotides /nucleoside 22
1.8.2 ADME genes associated with sensory neuropathy 26
1.8.3 ADME genes associated with CD4+ T cell recovery after ART 27
1.9 AIM AND OBJECTIVES 30
1.9.1 Aim 30
1.9.2 Objectives 31
CHAPTER 2: METHODS AND MATERIALS 33
2.1 Cohort 33
2.1.1 Patient recruitment 33

2.1.2 DNA Extraction and quantification 33
2.1.3 Phenotype 34
2.2 SNP selection 37
2.2.1 Literature review of ADME genes 37
2.2.2 MAF≥5% in the LWK population 38
2.2.3 Predicted functional effect 38
2.2.4 Tag SNPs 39
2.2.5 Sequenom software 39
2.3 Genotyping using Sequenom mass-array system / -iPLEX assay .. 40
2.3.1 Quality Control of Sequenom SNP data 46
2.4 qPCR assay for mtDNA copy number quantification 47
2.4.1 Optimization of the mtDNA copy number quantification assay 48
2.4.2 Optimization of qPCR 49
2.4.3 Assay validation 52
2.4.4 qPCR Plate designs for cohort analysis 53
2.5 Statistical analysis 55
2.5.1 Demographic data and SN 55
2.5.2 mtDNA copy number analysis 55
2.5.3 SNP analysis 55
CHAPTER 3: RESULTS 61
3.1 Demographic and clinical data and their association with sensory neuropathy and CD4+ T cell count 61
3.2 Cohort DNA Quantification 65
3.3 SNP Choice 65
3.4 Sequenom Results 69
3.4.1 Minor allele frequencies in the SA population 71
3.4.2 Minor allele frequencies in SA vs other African populations 74
3.4.3 LD and Haplotype analyses in the SA cohort. 77

3.4.4 SNP Genotype frequencies in SA population 81
3.5 Associations between SNPs and sensory neuropathy..... 82
3.5.1 Univariate SNP association with SN..... 82
3.5.2 Multivariate analysis of SNP association with SN..... 85
3.5.3 Analysis of associations between haplotypes and SN..... 88
3.6 Association between mtDNA copy number and sensory neuropathy..... 91
3.6.1 Primer optimization using regular PCR..... 91
3.6.2 qPCR optimization..... 93
3.6.3 Associations between mtDNA copy number and sensory neuropathy. 97
3.7 Associations between SNPs and mtDNA copy number..... 98
3.7.1 Univariate analysis for association with mtDNA copy number.. 99
3.8 SNP association with the CD4+ T cell count..... 102
3.8.1 Univariate analysis of association between SNPs and CD4+ T cell count 102
3.8.2 Multivariate linear regression analysis of associations between SNPs and CD4+ T cell count..... 106
3.8.3 Analysis of associations between haplotypes and CD4+ T cell count. 110
CHAPTER 4: DISCUSSION 114
4.1 Genetic variation in ADME genes in the black SA population..... 116
4.2. Demographic factors associated with phenotypes (SN, mtDNA copy number and CD4+ T cell count after ART). 117
4.3 Associations between SNPs and sensory neuropathy..... 118
4.4 Associations between mtDNA copy number and sensory neuropathy..... 121
4.5 Associations between SNPs and mtDNA copy number..... 122
4.6 Associations between SNPs and CD4 + T cell count 122
4.7. Strengths and weaknesses of the study 126
Conclusion..... 127
REFERENCES 128

APPENDIX A: ETHICAL CLEARANCE 152
APPENDIX B: GENOTYPE FREQUENCIES OF SA POPULATION. ... 153
APPENDIX C: MASS EXTEND EXTENSION PRIMERS. 156

LIST OF FIGURES

Figure 1.1	Diagram of phosphorylation of d4T and other NRTIs.	5
Figure 1.2	The proposed model by which NRTIs exert their toxicity to the mitochondria.	9
Figure 1.3	*De novo* synthesis of pyrimidines.	14
Figure 1.4	Synthesis of thymidylate.	16
Figure 1.5	Folate cycle-mediate dTMP synthesis	18
Figure 2.1	Patient exclusion diagram modified from Wadley et al. (2014).	36
Figure 2.2	iPLEX assay	44
Figure 3.1	An example of the results from Sequenom multiplex genotyping assay.	69
Figure 3.2	Example of mass spectrum for rs9905016 Heterozygote, with two peaks at 5330 and 5410 Da respectively.	69
Figure 3.3	Example of mass spectrum for rs9905016 homozygote with only one peak at 5330 Da.	70
Figure 3.4	*RRM1* gene LD plot generated using Haploview	77
Figure 3.5	*TK2* gene LD plot generated using Haploview	78
Figure 3.6	*SAMHD1* gene LD plot generated using Haploview	78
Figure 3.7	*SLC28A1* gene LD plot generated using Haploview	79
Figure 3.8	*ABCG2* gene LD plot generated using Haploview.	79
Figure 3.9	Agarose gel electrophoresis images of 1% (w/v) of the PCR products for *GAPDH* primer pairs.	91
Figure 3.10	The agarose gel electrophoresis images of 1% (w/v) of the PCR products for *mt-ND1* primer pairs.	92
Figure 3.11	Standard curve graph used to determine the PCR efficiencies of *GAPDH*.	93
Figure 3.12	Standard curve graph used to determine the PCR efficiencies of *mt-ND1* primers	93
Figure 3.13	A melting curve analysis test for *GAPDH* primer Pair 1.	95
Figure 3.14	A melting curve analysis test for *GAPDH* primer pair 2.	95
Figure 3.15	A melting curve analysis test for *mt-ND1* primer pair 2.	95
Figure 3.16	The comparison of mtDNA/nDNA (Ct) ratios between SN cases and controls, using the non-parametric Man-Whitney test.	98

LIST OF TABLES

Table 1.1.	South African National First line Antiretroviral Treatment Guidelines 2004-2016.	3
Table 1.2.	Enzymes which regulate the thymidine pool.	20
Table 1.3.	The drug transporters of d4T.	25
Table 2.1	iPLEX multiplex PCR and reagents	42
Table 2.2	IPLEX-SAP cycling conditions and recipe.	43
Table 2.3	Primer extension cycling conditions and recipe.	44
Table 2.4	PCR primers used for optimization of mtDNA copy number assay.	47
Table 2.5	PCR reaction mix.	48
Table 2.6	qPCR conditions for mtDNA quantification.	51
Table 2.7	Allelic and Genotypic models used in association tests.	58
Table 3.1	Associations between demographic and clinical factors with SN.	62
Table 3.2	Association of demographic data with mtDNA copy number.	63
Table 3.3	Associations between demographic and clinical variables with CD4+T cell counts at time of sample collection.	64
Table 3.4	SNPs of interest meeting our criteria for SNP selection, with 26 indicated by * selected for genotyping.	66-68
Table 3.5	Minor allele frequencies (MAFs) for d4T transporter genes and dNTP regulator genes in the SA cohort.	72-73
Table 3.6	Comparison of minor allele frequencies of the SA cohort with MAF of other African populations	75-76
Table 3.7	Haplotype frequencies in the SA cohort	80-81
Table 3.8	Univariate analyses for association of alleles and genotypes with sensory neuropathy.	83-84
Table 3.9	Multivariate analysis of SNP association with d4T-induced sensory neuropathy.	86-87

Table 3.10	Univariate and multivariate analysis of associations between haplotypes and sensory neuropathy.	89-90
Table 3.11	qPCR efficiencies for the primers selected for optimization.	94
Table 3.12	qPCR threshold cycle (Ct) values obtained using the DNA of HEK293 cells at three time points.	96
Table 3.13	The fold change ratios of *mt-ND1* relative to *GAPDH* in HEK293 cells.	97
Table 3.14	Fold change calculations for SN cases and controls.	97
Table 3.15	Univariate analyses for association with mtDNA copy number.	100-101
Table 3.16	Univariate analysis of associations between SNPs and CD4+ T cell count.	104-105
Table 3.17	Multivariate analysis of associations between SNPs and CD4+ T cell count.	108-109
Table 3.18	Univariate and multivariate analysis of associations between haplotypes and CD4+ T cell count.	111-112

LIST OF ABBREVIATIONS

°C	degrees Celsius
3TC	Lamivudine
∞	Infinity symbol
ABCC	ATP Binding Cassette Subfamily C Member
ABCG	ATP-binding cassette transporters
ACTG	AIDS Clinical Trials Group
ADME	"absorption, distribution, metabolism, and excretion,"
AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Therapy
ARV	Antiretroviral
ATN	Antiretroviral toxic neuropathy
ATN	ART-induced Neuropathy
AZT	Azidothymidine
BMI	Body Mass Index
bp	Base pair
cART	Combination Antiretroviral Therapy
cm	Centimetre
C_{max}	Maximum Drug Plasma Concentration
CNT	Concentrative nucleoside transporter
CO_2	Carbon dioxide
d4T	2′,3′-didehydro-2′,3′-dideoxythymidine (Stavudine)
d4T-DP	stavudine disphosphate
d4T-MP	stavudine monophosphate

d4T-TP	stavudine trisphosphate
Da	Dalton
ddC	Zalcitabine
ddI	Didanosine
DHFR	Dihydrofolate Reductase
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotide Triphosphates
dsDNA	Double stranded DNA
DSP	Distal Sensory Polyneuropathy
dTMP	deoxythymidine monophosphate
dTTP	Deoxythymidine triphosphate
EDTA	Ethylenediaminetetraacetic acid
EFV	Efavirenz
EtBr	Ethidium bromide
FTC	Emtricitabine
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Virus
HIV-SN	HIV-associated sensory neuropathy
HWE	Hardy Weinberg Equilibrium
IBD	Identity By Descent
INRs	Immunological non-responders
Kb	Kilobase
kg	kilogram
LD	Linkage disequilibrium

LWK	Luhya in Webuye, Kenya
MAF	Minor allele frequency
MHC	Major Histocompatibility Complex
mtDNA	Mitochondrial Deoxyribonucleic Acid
MTHFR	Methylenetetrahydrofolate Reductase
MT-ND1	Mitochondrial encoded NADH-ubiquinone oxidoreductase chain 1
NaCl	Sodium Chloride
nDNA	Nuclear DNA
ng	Nanogram
NNRTI	Non-reverse Nucleoside Reverse Transcriptase Inhibitors
NRTI	Nucleoside Reverse transcriptase inhibitor
NTC	No template control
NVP	Nevirapine
PCR	Polymerase Chain Reaction
P_{EMP1}	pointwise empirical p-value
P_{EMP2}	Family-wise correction for multiple comparisons
PN	Peripheral neuropathy
POLG	DNA polymerase gamma
qPCR	quantitative PCR
RNA	Ribonucleic Acid
RNR	Ribonucleoside reductase
ROS	Reactive Oxidative Species
RRM1	Ribonucleotide Reductase Catalytic Subunit M1
RRM2	Ribonucleoside-diphosphate reductase subunit M2
RRM2B	Ribonucleotide Reductase Regulatory TP53 Inducible Subunit M2B

SAMHD1	SAM domain and HD domain-containing protein 1
SAP	shrimp alkaline phosphatase
SBE	Single base extension
SDS	Sodium lauryl sulfate/Sodium dodecyl sulphate
SN	Sensory neuropathy
SNP	Single nucleotide polymorphism
ssDNA	Single stranded DNA
SSPs	Single specific primers
TBE	Tris-borate-EDTA
TDF	Tenofovir
TK	Thymidine Kinase
TS	Thymidylate synthase
V	Volts
VNTR	Variable Number Tandem Repeat
YRI	Yoruba in Ibadan, Nigeria
µg	microgram
µl	microlitre
µM	micromolar

CONFERENCE PRESENTATIONS ARISING FROM THIS STUDY

Marutha, TR., **Moketla, MBM.,** Wadley, A., Kamerman, P., de Assis Rosa, D., 2015. SNP selection in candidate genes: Towards Pharmacogenetic studies of ART-related sensory neuropathy. Poster Presentation at the Southern African Society of Human Genetics (SASHG) conference in Pretoria, South Africa [16-19 August 2015].

Moketla, MBM., Wadley, A., and Kamerman, P., de Assis Rosa, D.,2016. Variation in the *SLC28A1* transporter gene is associated with sensory neuropathy in a South African HIV+ cohort exposed to d4T. Poster presentation at the Molecular Biosciences Research Thrust (MBRT) Postgraduate Research day 2016, at the Wits club, Johannesburg, South Africa [08 December 2016].

CHAPTER 1

INTRODUCTION

CHAPTER 1: INTRODUCTION

1.1 HIV and HAART

Human Immunodeficiency Virus (HIV) is still a major health threat with over 36.9 million people living with HIV, of which 25.8 million are based in sub-Saharan Africa accounting for almost 70% of all people living with HIV by the end of 2015 (Joint United National Global Fact Sheet, 2015). However, the introduction of antiretroviral drugs (ARVs) since 1995 (2003 in South Africa) has caused a decline in the mortality and morbidity rate of those living with HIV/AIDS and accordingly the life expectancy of HIV/AIDS patients has improved (Brinkman et al., 1999; Menezes et al., 2013). There are six classes of ARVs, including nucleoside reverse transcriptase inhibitors (NRTIs), non-reverse nucleoside transcriptase inhibitors (NNRTIs), protease inhibitors (PI), integrase inhibitors (INSTI), fusion inhibitors, and co-receptor antagonists (Pirmohamed & Back, 2001). A three drug combination referred to as highly active antiretroviral therapy (HAART) was introduced worldwide in 1996 as a standard course of therapy (Brinkman et al., 1999). The NRTIs serve as the backbone of virtually all the ART regimens which usually comprise of two NRTIs plus a NNRTI, a PI or INSTI (Marcelin et al., 2014). In South Africa (SA) HAART was initiated in 2003 with the SA Government ART roll-out commencing in 2004 according to the national ART programme (Menezes et al., 2013; National Antiretroviral Treatment guidelines SA, 2004; National Consolidated ART Guidelines SA, 2016, **Table 1.1**).

Table 1.1 South African National First line Antiretroviral Treatment Guidelines 2004-2016.

Year	Regimen 1a	CD4+ T cell count at treatment initiation
2004	d4T+ 3TC + EFV	Starting at CD4+ T cell counts < 200 cell/mm^3
2010	TDF+ 3TC+ EFV	Starting at CD4+ T cell counts < 350 cell/mm^3
2015	TDF+ (FTC or 3TC)+ EFV	Starting at CD4+ T cell counts < 500 cell/mm^3
2016	TDF+ (FTC or 3TC)+ EFV	Starting at any CD4+ T cell count

Legend: d4T (stavudine), 3TC (lamivudine), NVP (nevirapine), TDF (tenofovir), EFV (efavirenz), FTC (Emtricitabine) and cell/mm^3 (cells per millimeter).

Stavudine-based regimens were the baseline of the South African ART programme for approximately six years as part of the first-line public sector HAART, but due to the adverse effects of stavudine, since 2010 the South African ART guidelines replaced stavudine with tenofovir (TDF) (Menezes et al., 2013).

The use of ART usually successfully reduces the HIV viral load by halting its replication in many individuals and thus boosts CD4+ T-cell recovery which lead to the decrease of AIDS associated deaths and diseases (Baker et al., 2008). It has been noted that 15-30% of the patients on long-term ART do not respond to the treatment which can result in the lack of CD4+ T-cell recovery to normal levels amidst full suppression of HIV replication; these patients are referred to as immunological non-responders (INRs) (Gazzola et al., 2009). Several clinical factors have been associated with poor CD4+ T cell recovery after ART, including baseline CD4+ T cell count, thymus function, anaemia, age, advanced stage of disease, and co-infections (Hunt et al., 2003; Aiuti and Mezzarroma, 2006; Baker et al., 2008; Muzah et al., 2012; Massanella et al., 2013). Immunological non-response, or varying levels and rates of CD4+ T cell recovery, may also be influenced by genetic factors (e.g. Parathyras et al., 2009).

1.2 Nucleoside reverse transcriptase inhibitor (NRTIs)

HIV-1 reverse transcriptase (RT) was discovered in the early stages as a potential target for antiretroviral treatment. HIV-1 RT is responsible for synthesis of a double-stranded proviral DNA from the viral RNA (Menendez-Arias, 2008). Nucleoside reverse transcriptase inhibitors (NRTIs) are analogues of the deoxyribonucleosides but lack the 3'-hydroxyl group (OH) on their sugar moiety (Sluis-Cremer and Gilda Tachedjian, 2008). The NRTI analogues compete with the natural deoxynucleotide triphosphates (dNTPs) for incorporation into the newly synthesized strand, the inclusion and/or insertion of these nucleoside analogues results in viral DNA chain termination (de Béthune, 2010; Gallant, 2002; Sluis-Cremer et al., 2004; Kakuda, 2000).

There are eight NRTIs currently approved for clinical use, namely stavudine (d4T), lamivudine (3TC), didanosine (ddI), tenofovir (TDF), Abacavir (ABC), emtricitabine (FTC), Zidovudine (AZT), Zalcitabine (ddC) (Cihlar and Ray, 2010).

The NRTIs in their administered form are inactive and require phosphorylation by cellular kinases to form active deoxynucleoside triphosphates (dNTP) analogues capable of viral inhibition (Kakuda, 2000; Figure 1.1). The antiviral activity of NRTIs is significantly influenced by intracellular phosphorylation kinetics (Stein and Moore, 2001). The sequential steps of NRTI phosphorylation from monophosphate to diphosphate to triphosphate can function as a rate-limiting step in the formation of an active drug (Stein and Moore, 2001). The intracellular concentration of NRTI-TP is regulated by a feedback mechanism of one or more enzymes in the phosphorylation pathway (Stein and Moore, 2001).

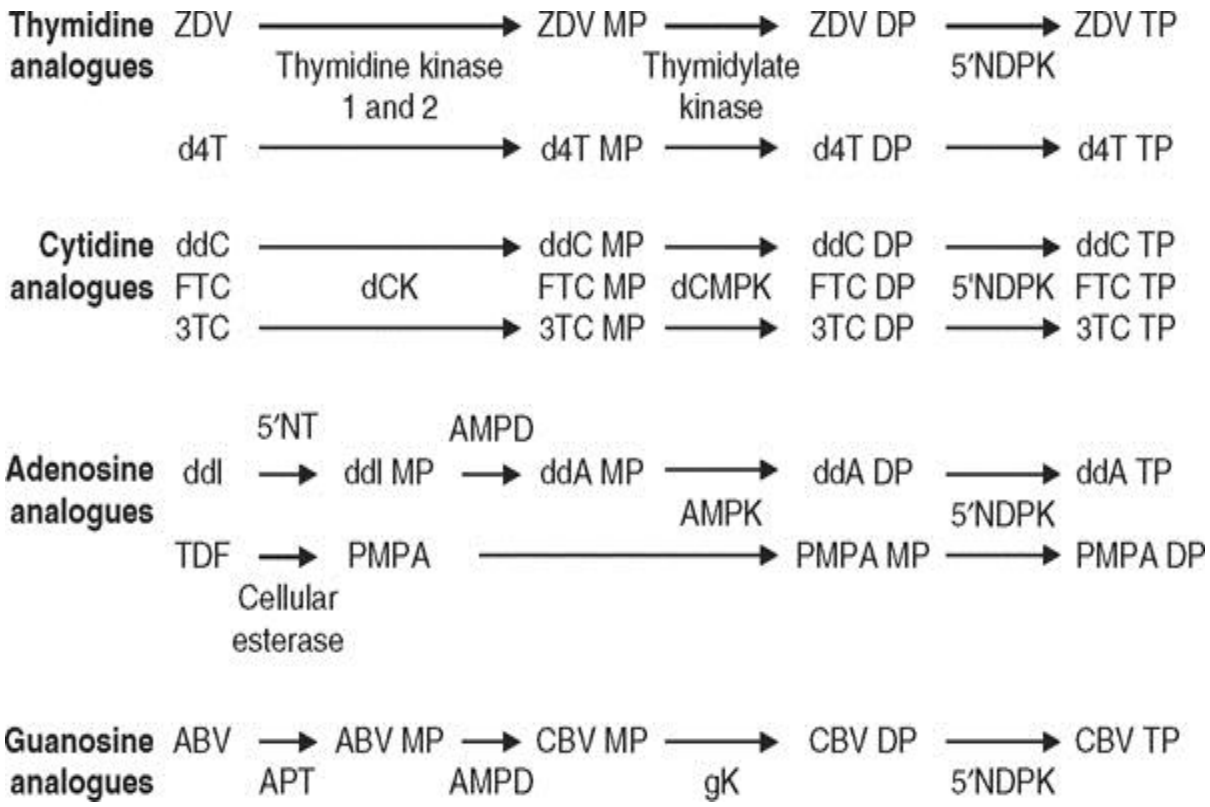

Figure 1.1 The serial phosphorylation of NRTIs from their inactive administered form (prodrug) to their active triphosphate form (NRTI-TP), ABV, abacavir; d4T, stavudine; AMPD, adenosine monophosphate deaminase; AMPK, adenosine monophosphate kinase (Adenylate kinase); APT,adenosine phosphotransferase; dCMPK, deoxycytidine monophosphate kinase; ddA , 2',3'-dideoxyadenosnie; ddI, didanosine; ddc, 2′-3′-dideoxycytidine; DP, diphosphate; CBV, carbovir; dCK, deoxycytidine kinase; ddC, Zalcitabine; FTC, emtricitabine; gK, guanylate kinase MP, monophosphate; PMPA , tenofovir (PMPA DP is a triphosphate analogue); TDF, tenofovir disoproxil fumarate; TP, triphosphate; ZDV, zidovudine; 3TC,lamivudine; 5'NDPK, 5'nucleoside diphosphate kinase; 5'NT,5'nucleotidase (Anderson et al., 2004).

1.3 Stavudine

Zerit (Stavudine, d4T), 2'3'-didehydro-3' deoxythymidine is a thymidine analogue that acts as a chain terminator like other NRTIs (Piliero, 2004). Stavudine is administered orally to HIV patients, and its oral bioavailability ranges from 82-99% for the dose range 0.1 to 12 mg/kg (Lea and Faulds, 1996). Stavudine is reported to have an elimination half-life of 1.5 hours for a standard dose of 40mg in HIV infected patients and a corresponding half-life of 3.5 hours for the intracellular active metabolite stavudine-5'-triphosphate, in the peripheral blood mononuclear cells (Lea and Faulds,1996). Approximately 40% of stavudine in the plasma is renally excreted through the glomerular filtration and tubular secretion, with the remaining percentage thought to be eliminated through metabolism of pyrimidine salvage pathways and possibly degradation (Dudley et al., 1992; Hurst and Noble, 1999; Margolis et al., 2014).

The peak plasma concentration (C_{max}) and area under the concentration-time curve (AUC) of stavudine occurred at a duration (T_{max}) 0.50 to 0.75 hours after oral dosing following a dose-dependent manner (Lea and Faulds, 1996). The peak plasma concentration of stavudine is affected by a high fat meal, which also delays the time to reach C_{max}, but the overall bioavailability of stavudine is not affected (Riddler et al., 1995; Lea and Faulds, 1996). Plasma clearance of stavudine is not dose-dependent nor body weight dependent (Riddler et al., 1995).

1.4 Mitochondrial toxicity of NRTIs

Regardless of the positive effects of the HAART, exposure to HAART can yield severe adverse events and morbidities (Walker et al., 2002; Calmy et al., 2007). There are various complications associated with the use of ART in HIV positive patients. In particular, the NRTIs are associated with mitochondrial toxicity which results in a serious side effects such as hepatic steatosis (1% of the patients), peripheral neuropathy (7.5%), lipoatrophy (50%), lipomatosis (10%), and less frequently pancreatitis, cardiomyopathy, proximal myopathy, renal tubular acidosis (all<1%)(Carr, 2003).

Mitochondria are semi-independent organelles found in eukaryotic cells, which harbour their own genome and play a critical role in cellular energy production, formation of free radicals and apoptosis (Yu et al., 2013). The mitochondria harbour a diverse mtDNA genome that is 16 569 base pairs long and codes for 13 proteins as well as 2 ribosomal RNAs and 22 transfer RNAs that are signification for the translation of intra-mitochondrial proteins (Montier et al., 2009; Andreu et al., 2009).

Mammalian cells can have over 100 mitochondria per cell with some having as much as 1000 and furthermore each of these organelles has an mtDNA copy number ranging from two to ten, therefore a single human cell can have over 1000 mtDNA copies with some cells reaching 5000 mtDNA molecules (Moraes, 2001). Alterations in the nuclear genes involved in the synthesis of the mtDNA or responsible for the sustenance of deoxyribonucleotide molecules can greatly influence the mitochondrial copy number due to the inefficiency of mitochondrial biogenesis (Andreu et al., 2009).

NRTI drugs are associated with mitochondrial toxicity due to their inhibitory activity of DNA polymerase gamma (*POLG*), an enzyme responsible for mtDNA replication (Walker et al., 2002; Figure 1.2). This results in the damage and loss of the mtDNA including decreased mtDNA copy number in many tissues (Cherry et al., 2002). This consequently weakens the ability of mtDNA to synthesize oxidative enzymes responsible for energy production in the oxidative phosphorylation pathway (Carr and Cooper, 2000). There are significant differences in the relative potencies of NRTI in their ability to interact with *POLG*. The hierarchy of gamma-polymerase inhibition for the active NRTI metabolites has been determined as follows: zalcitabine (ddC,) > didanosine (ddI), > stavudine (d4T) > lamivudine (3TC) ≥ abacavir (ABC) ≥ tenofovir (TDF) ≥ emtricitabine (FTC) (Venhoff and Walker, 2011).

The effects of stavudine on the mitochondria manifests *in vivo* as a reduction in the mtDNA copy number per cell in tissues such as the human adipose tissue and blood and as a result cause associated mitochondrial toxicities including pancreatitis, SN, lipodystrophy (Cherry et al., 2002).

The toxicities attributed to d4T are believed to be established through the inhibition of the human mtDNA polymerase gamma which plays a critical role in the replication and repair of the mtDNA, this inhibition subsequently results in the depletion of the mtDNA (Lewis, 2003). NRTIs affect the functionality of the DNA polymerase gamma (*POLG*) which consequently perturbs the expression of vital mtDNA genes coding for proteins responsible for energy production and regulation of reactive oxygen species (ROS) (Kamerman et al., 2012).

In addition, changes in mtDNA copy number due to *POLG* inhibition may not be the only mechanism underlying antiretroviral toxic neuropathy (ATN). ATN is a form of SN associated with the use of certain nucleoside antiretroviral drugs, see section 1.5 (Lewis, 2003). Changes in the levels of dNTP pools can also be associated with ART-induced mitochondrial toxicity, independent of pol-γ inhibition (Selvaraj et al. 2014). Variation in the activity of d4T- specific drug transporters influences drug availability inside cells, d4T uptake by *POLG*, mtDNA copy number and resulting toxic side effects.

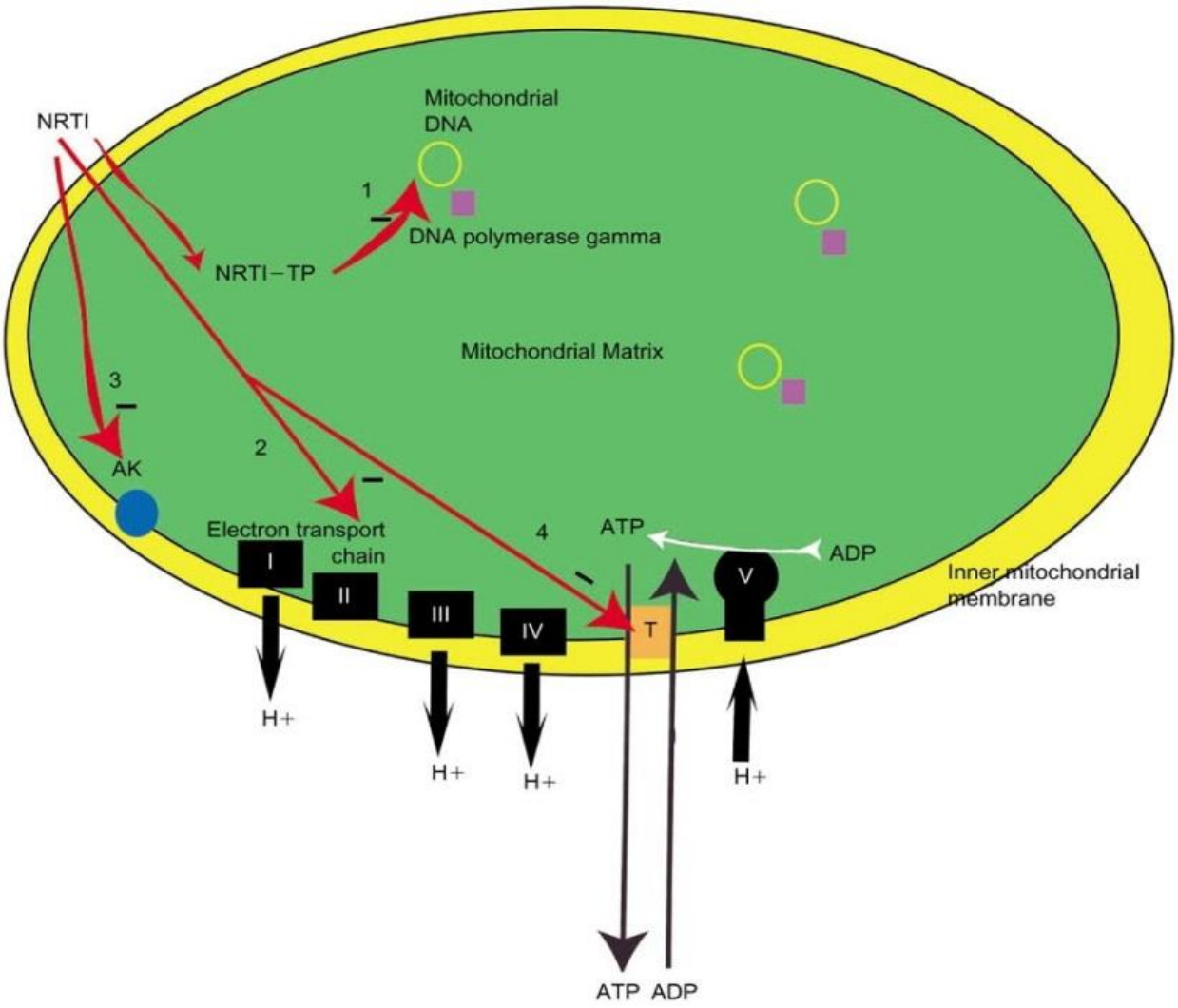

Figure 1.2 The proposed model by which NRTIs exert their toxicity to the mitochondria. The effects of the NRTIs (D4T) includes (1) the inhibition of the mtDNA polymerase gamma which consequently leads to the depletion of the mtDNA as well as the degradation of the oxidative phosphorylation pathway which is manifested in three ways comprising (2) the perturbation of the respiratory complexes involved in the formation of the electron transport chain (3) the disruption in the adenylate kinase (AK) function (4) and lastly the interference in the activity of the ADP/ATP translocator (T) (Keswani et al., 2002).

1.5 Sensory neuropathy

Sensory neuropathy (SN) is a neurological disorder characteristic of HIV infection but the prevalence of SN has increased significantly by the initiation of antiretroviral therapy (Cherry et al., 2003). HIV-SN is subdivided into two categories, distal sensory polyneuropathy (DSP) due to HIV infection, and the antiretroviral toxic neuropathy (ATN) which stems from the use of NRTIs (Keswani et al., 2002; Kamerman et al., 2012). The only distinguishing factor of the aetiology of ART-induced SN from DSP is the time of onset after the exposure of NRTIs (McArthur et al., 2005). Hao (2013) reported that DSP and drug-induced SN portray the same clinical manifestations. The symptomatic features of SN are predominantly characterised by pain, loss of sensation (McArthur et al., 2005), numbness, pins and needles and signs that can be measured including loss of ankle reflexes and loss of vibration sense (Wadley et al., 2011).

Both DSP and ATN affect the peripheral nerves (Pardo et al., 2001) and are characterised by the erosion of the size-dependent nerve fibres (Kamerman et al., 2012). The most common neurological descriptive feature of HIV-SN is the loss of dorsal root ganglions as well as the degeneration of long axons on the distal regions (Pardo et al., 2001), this phenomenon is often referred to as the ‘dying back’ pattern of axonal degeneration (Keswani et al., 2002).

The primary step after excluding metabolic causes in treating HIV-SN is the consideration for withdrawal and dose reduction of neurotoxic antiretroviral drugs such as didanosine, stavudine, zalcitabine (Wulff et al., 2000).

1.6 Pharmacogenetics of SN

Not all patients who are exposed to NRTIs develop mitochondrial toxicity such as sensory neuropathy. This suggests that factors such as underlying genetic variation can influence the development of SN in response to NRTI use. Pharmacogenetics refers to the study of altered drug response as a result of the genetic changes in the genes affecting the pharmacokinetics (drug transporters, drug metabolizers) and pharmacodynamics (drug targets) of drugs (Kalow, 2002;Johnson, 2003) and this may help to predict who suffers off-target side-effects of ART and who does not.
In particular, human genetic variation could influence SN development in a variety of ways:

1. Variation in the genes affecting drug transport into and out of cells could influence the relative amounts and concentrations of drug in plasma and inside cells of the tissues.
2. Variation in the genes affecting drug metabolism e.g. phosphorylation, activation, breakdown and excretion of the drug, further influencing relative amounts and concentrations of active drug in plasma and inside cells of the tissues.
3. Variation in genes affecting immune responses to HIV, including cytokines and the MHC region.
4. Variation in genes affecting pain neurology or pain perception.
5. Variation in genes controlling mtDNA copy number or function, since *POLG* inhibition is thought to be the mechanism by which NRTI toxicity arises.

Genes in (3) and (4) have been the focus of series of studies in the same cohort studied here, conducted by our collaborators in the School of Physiology, Faculty of Health Science, University of the Witwatersrand. Genes that have already been examined in the cohort under current investigation, include TNFa haplotypes, several cytokines, Uncoupling protein (*UCP*), GTP cyclohydrolase 1 (*GCH1*), Potassium voltage-gated channel subfamily 1 (*KCNS1*), Interleukin 1b (*IL1B*), Interleukin 6 (*IL6*), Monocyte chemoattractant protein-1 (*CCL2*) and CC motif chemokine receptor–2 (*CCR2*) (Hendry, 2014; Wadley et al., 2011; Wadley et al., 2013a; Wadley et al., 2013b; Wadley et al.,

2014). Genes in (5) are the subject of another current study in the School of Molecular and Cell Biology.

Collectively the pharmacogenes in (1) and (2) are sometimes referred to as the absorption, distribution, metabolism, and excretion (ADME) genes since they determine drug absorption, distribution, metabolism and excretion / elimination (Kerb et al., 2009). These can be core determinant of the toxicities and efficacies of ARVs (Michaud et al. 2012; Johnson, 2003). In addition, many of the same ADME genes also control the relative levels of the natural dNTP pool with which the NRTIs compete inside the cells. Variation in endogenous dNTP or RN levels could be a key second mechanism by which NRTI toxicity occurs (Selvaraj et al., 2014). Decreased endogenous dNTP pools result in higher incorporation of the NRTI triphosphate (Rampazzo et al. 2010) and could cause both increased NRTI efficacy and increased NRTI toxicity. For example, Selvaraj et al. (2014) found that HIV infected individuals on ART who had mitochondrial toxicity had lower ribonucleotide (RN) and deoxyribonucleotide (dRN) pools compared to infected individuals on ART who did not have mitochondrial toxicity.

Genetic variations in the ADME genes and their influence on the development of SN and on CD4 count recovery are the focus of the current study. The following sections therefore review: Thymidine/dTMP synthesis and metabolism pathways (in section 1.7) as well as dTTP and d4T drug transporters (in section 1.8).

1.7 dNTP synthesis and metabolism pathways

The cell requires the four dNTPs (adenosine, thymidine, cytosine, and guanine) in adequate supply in order to replicate and repair its DNA efficiently, for sufficient mitochondrial function and in order for cell division to occur (Reichard, 1988). The availability of dNTP in cells is governed by two primary pathways, namely, the *de novo* pathway and the salvage pathway (Rompay et al., 2000).

1.7.1 *de novo* synthesis of pyrimidines

The liver is the major organ of *de novo* synthesis of all four nucleotides. The *de novo* dNTP pathway involves creation of new dNTP precursors (deoxyribonucleotides) from carbohydrate and amino acid metabolism (Christiansen et al., 2015). *De novo* synthesis is cell cycle-regulated, operative during the S-phase of DNA replication in cells (Christiansen et al., 2015).

The *de novo* synthesis of all pyrimidines occurs by the following steps:

- Glutamine and bicarbonate are combined to make carbamoyl-phosphate.
- Carbamoyl-phosphate is then used to make the common precursor ring structure orotic acid (orotate) (O'Donovan & Neuhard, 1970).
- A phosphorylated ribosyl unit (phosphoribosylpyrophosphate or PRPP) is added to orotic acid, to create a pyrimidine nucleotide called orotidylate (OMP) (O'Donovan & Neuhard, 1970).
- Decarboxylation of OMP yields uridine monophosphate (uridylate or UMP).
- It is from UMP that other pyrimidine nucleotides are derived, including both ribonucleotides and deoxyribonucleotides (Löffler et al., 2005).

  All the steps summarised above are illustrated in the figure below (Figure 1.3).

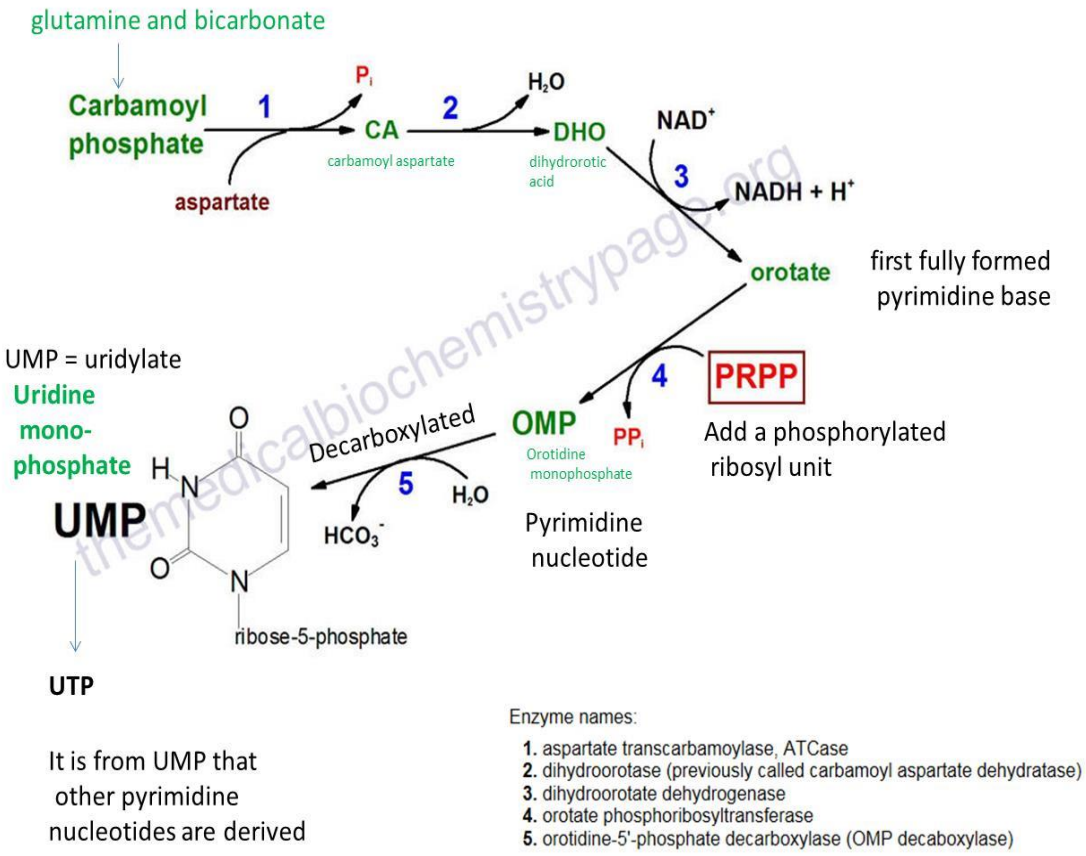

Figure 1.3 *de novo* synthesis of pyrimidines (modified from www.themedicalbiochemistrypage.org, 2015). This process is explained further below.

This *de novo* pathway is used for synthesis of both the ribonucleotides and the deoxyribonucleotides.

However due to relatively large pools of ribonucleotides compared to deoxyribonucleotides, the major pathway of deoxyribonucleotide synthesis is from existing ribonucleotides being converted to deoxyribonucleotide (e.g. UMP to dTMP) by the enzyme ribonucleotide reductase (RNR).

The RNR is the only enzyme capable of converting ribonucleotide to deoxyribonucleotides and therefore is a rate-limiting step for the production of 2-deoxyribonucleoside 5-triphosphates (dNTPs) that are necessary for DNA replication (Heidel et al., 2007). Ribonucleotide reductase (RNR) is a multimeric enzyme that consists of two subunits: the large subunit (RRM1 or R1) and the small subunit (RRM2 or R2) (Jørgensen et al., 2013).

The R1 protein is present throughout the cell cycle while the R2 subunit is only expressed during the late G and early S phase during DNA replication (Heidel, 2007).

For purposes of this study the synthesis of dTTP, the competitor of d4T, is described, and shown in Fig 1.4. UMP is phosphorylated to UDP by uridylate kinase and UDP is then converted to dUDP by the enzyme ribonucleotide reductase (RNR) (Hauschka, 1973). dUDP does not get converted directly to dUMP. dUDP is phosphorylated again to dUTP by nucleoside diphosphate kinase, which is then broken down to dUMP by dUTPase.

Finally, dUMP is converted to thymidine monophosphate (dTMP or thymidylate) by the enzyme thymidylate synthase. dTMP can then be further phosphorylated by the enzymes thymidylate kinase and nucleoside diphosphate kinase to dTTP, the active form of thymidine used in DNA synthesis (Christiansen et al., 2015).

Isoforms of some of the enzymes in the de novo pathway are found only in the mitochondria and allow ongoing dNTP synthesis in non-dividing cells (Christiansen et al., 2015), including:

- P53R2 or RRM2B isoform of RNR
- mitochondrial isozyme of DHFR called DHFRL1
- mitochondrial form of dUTPase

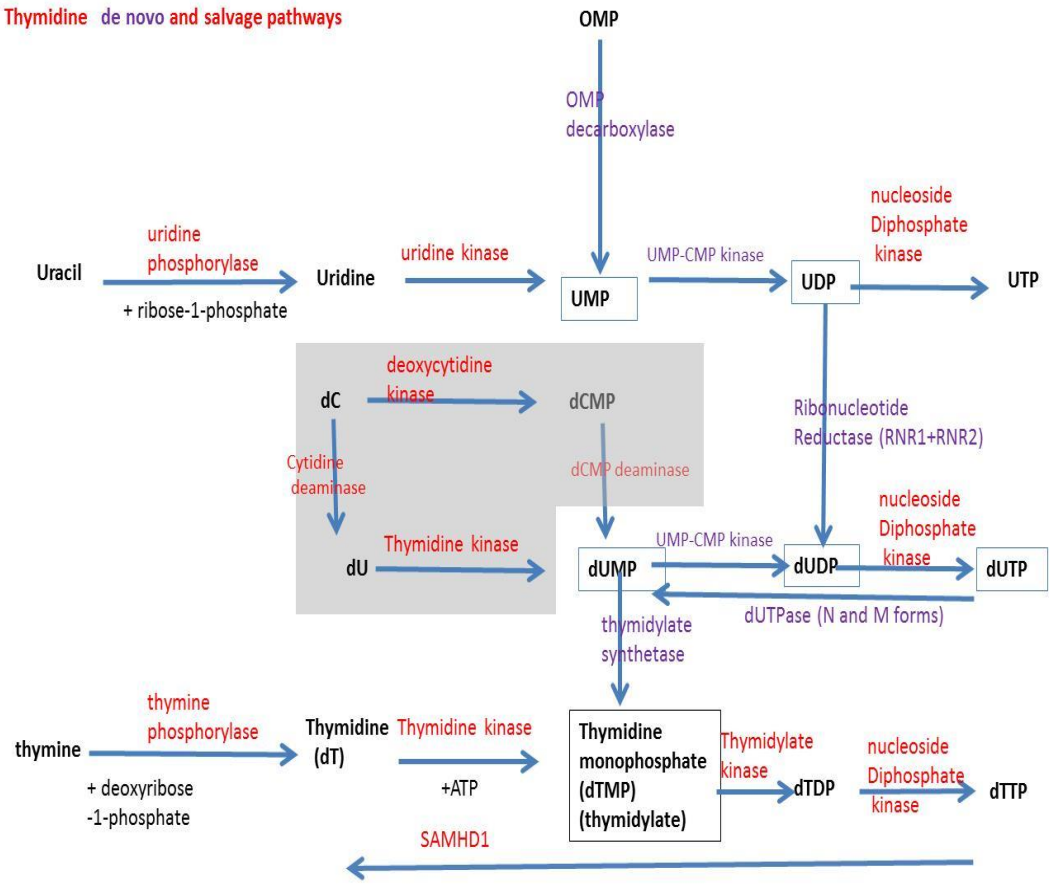

Figure 1.4 Thymidylate synthesis pathways (collated by D de Assis Rosa). Substances in boxes/purple texts are part of the *de novo* synthesis pathway, while all other substances are part of the salvage pathway. OMP, Orotidine 5'-phosphate; UMP, Uridine 5'-monophosphate; UDP, Uridine 5'-diphosphate; UTP, Uridine-5'-triphosphate, dC, 2'-deoxycytidine; dU, 2'-deoxyuridine; dCMP, 2'-Deoxycytidine 5'-monophosphate; dUMP; 2'-deoxyuridine-5'-monophosphate;dUDP, 2'-Deoxyuridine 5'-diphosphate;dUTP, 2'-Deoxyuridine 5'-triphosphate;dTTP, 2'-deoxythymidine-5'-triphosphate;dTDP, thymidine 5'-diphosphate ;dTMP, Thymidine Monophosphate ; ATP, Adenosine triphosphate ;SAMDH, The sterile alpha motif and HD-domain containing protein 1.

1.7.2 Folate cycle

The production of dTMP by thymidylate synthase (TS) is tightly linked to the folate / methionine cycle **(Figure 1.5.).** Folates are naturally occurring water-soluble members of the Vitamin-B class essential for maintaining the synthesis of thymidine and purines, actively involved in the amino acids conversions and also required for the methylation of essential biological substances such as DNA, proteins, phospholipids and neurotransmitters (Chango et al., 2000; Hamid et al., 2009; Crider et al., 2012).

Dihydrofolate reductase (DHFR) enters the system from dietary folic acid (vitamin B9). The reduced folate carrier RFC encodes a transmembrane anionic exchanger which serves as carrier of folate into the cells. The enzyme DHFR converts the DHF to tetrahydrofolate (THF) (Main et al., 2010). THF becomes a recipient of a methyl group during the conversion of serine to glycine catalysed by enzyme SHMT, and THF is converted to (5, 10-MTHF) (de Silva & Davis, 2013). During *de novo* dTMP synthesis, enzyme thymidylate synthase (TS) uses a methyl group donated by 5, 10 methylene tetrahydrofolate (5, 10-MTHF) as a cofactor to maintain the dTMP pool necessary for DNA repair and synthesis (Trinh et al., 2002) (**Figure 1.5**).

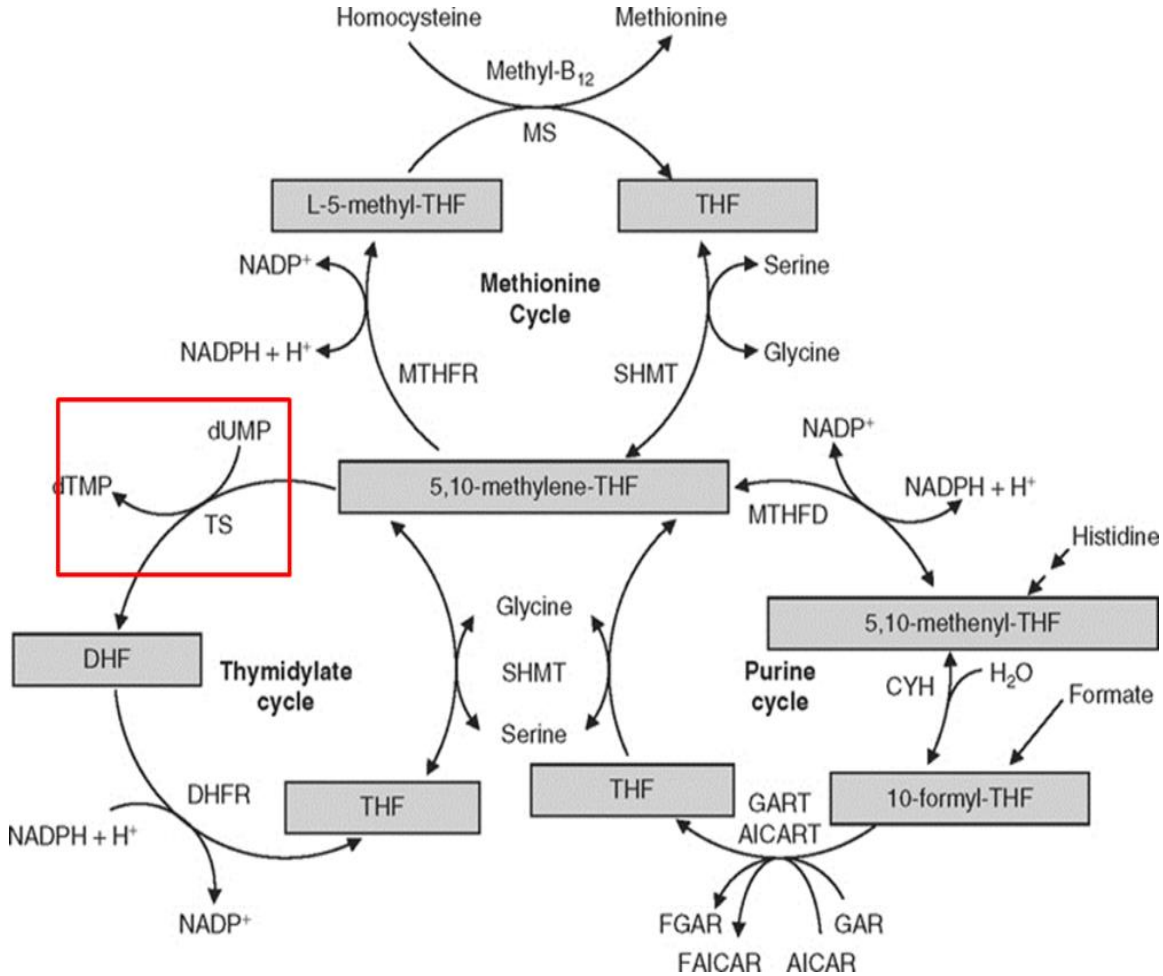

Figure 1.5 Folate and the thymidylate cycle mediated dTMP synthesis. **Red box indicates occurrence of dTMP production within this cycle.** AICAR= aminomidazole carboxamide ribonucleotide; AICART = AICAR formyl transferase; B=vitamin B (cobal-amin);CYH = methenyl-THF-cyclohydrolase; DHF= dihydrofolate; DHFR=DHF reductase; dTMP = deoxythymidine monophosphate; dUMP= Deoxyuridine monophosphate; FAICAR = formyl-AICAR;FGAR =formyl-GAR; GAR = glycinamide ribonucleotide; GART=GAR formyl transferase; MS=Methionine synthase; MTHFD = methylene-THF-dehydrogenase; MTHFR=5,10-methylene-THF-reductase;NADP = nicotinamide adenine dinucleotide phosphate; NADPH = reduced NADP; SHMT = serine hydroxyl methyltransferase ; THF=tetrahydrofolate; TS = thymidylate synthase (Pietrzik et al., 2010).

1.7.3 Salvage pathway of dNTP production

The salvage pathway is a complementary process for dNTP synthesis (Figure 1.3). The activation of a number of antiviral prodrugs (including d4T) is also performed using the salvage pathway (Sandrini & Piskur, 2005). The phosphorylation of deoxyribonucleosides to their monophosphate metabolites is catalysed by deoxyribonucleoside kinases (Knecht et al., 2002). The two deoxyribonucleoside kinases relevant to thymidine salvage and d4T activation in humans are thymidine kinase 1 (TK1, active in cytosol) and thymidine kinase 2 (TK2, active in mitochondria). There are also two deoxyribonucleoside kinases not relevant to thymidine salvage or d4T activation: deoxycytidine kinase (dCK) and deoxyguanosine kinase (dGK), due to the fact that these two kinases do not play a role in the phosphorylation of d4T to the active metabolite d4T-TP, which begins with the rate-limiting step of *TK1* or *TK2* to d4T-MP.

Lastly, dTTP may be converted back to unphosphorylated thymine (dT) by the enzyme SAMHD1; however, this enzyme does not hydrolyse active d4T back to its unphosphorylated form (Ballana et al. 2014). *SAMHD1* activity may increase efficacy of thymidine analogs by reducing competition with intracellular dNTP (Amie et al., 2013). dNTP synthesis by the salvage pathway, particularly in the presence of mitochondrial damage, may be boosted by the activity of PINK1, a protein more traditionally thought to be involved in Parkinson's disease (Tufi et al. 2014).

In summary, the following **Table 1.2** shows the key enzymes which influence *de novo* and salvage thymidine synthesis and therefore control endogenous levels of dTMP and its derivatives such as dTTP. Many of these enzymes play a role in the phosphorylation, activation and metabolism of both endogenous thymidine and d4T.

Table 1.2 Enzymes which control thymidine and d4T metabolism.

Enzyme	Abbreviation	Gene	Chromosomal position of gene
Ribonucleotide reductase subunit 1	RNR1	*RRM1*	11p15.5
Ribonucleotide reductase subunit 2	RNR2	*RRM2*	2p25-p24
Ribonucleotide Reductase Regulatory TP53 Inducible Subunit M2B	p53R2	*RRM2b*	8q22.3
Nucleoside Diphosphate Kinase	NDPK-A	*NME1*	17q21.3
deoxyuridine triphosphatase	dUTPase	*DUT*	15q21.1
Thymidylate synthetase	TS	*TYMS*	18p11.32
Dihydrofolate reductase	DHFR	*DHFR*	5q11.2-13.2
Dihydrofolate reductase-like 1	DHFRL1	*DHFRL1*	3q11.1
Methylenetetrahydrofolate	MTHF	*MTHFD1*	14q24
Thymidine kinase	TK1	*TK1*	17q23.2-q25.3
Mitochondrial thymidine kinase 2	TK2	*TK2*	16q21
Thymidylate Kinase	dTMP kinase	DTYMK	2q37.3
SAM domain and HD domain 1	SAMHD1	*SAMHD1*	20pter-q12
PTEN induced putative kinase 1	PINK1	*PINK1* *PARK6*	1p36

1.8 Transporter proteins

Membrane transporter proteins are responsible for the influx and efflux of critical compounds across biological membranes (Hediger et al., 2004; Goole et al., 2010). The transporters expressed in tissues like the intestine, liver, brain and kidneys play a very crucial role in the absorption, distribution and excretion of many drugs used daily in clinical practise (Telenti and Zanger, 2008). There are two types of drug transporters: uptake and efflux transporter proteins. The uptake transporters facilitate the influx of biological compounds across the lipid bilayer in the cytosol, whereas efflux transporters mediate the efflux of compounds out of the cells (Hediger et al., 2004; Goole et al., 2010).

Transporter proteins are clustered into two major families of transporters, namely the solute carriers (SLC) and the ATP-binding cassette (ABC) superfamilies (Estudante et al., 2013) which are generally uptake and efflux transporter proteins respectively (Konig et al., 2013). The SLC proteins facilitate movement of substances across the membranes into cells through the use of the electro-chemical gradient of solutes, whereas the ABC proteins utilize the energy derived from the hydrolysis of ATP in order to move substrates across the membranes out of cells as well as to circumvent the concentration gradients of the solutes (Cascorbi, 2006; De Gorter et al., 2012).

The SLC superfamily of transporter proteins comprises of more than 300 proteins which are classified into subfamilies. The subfamilies are assigned numbers from 1 to 52 after the root name SLC and the corresponding genes are identified by the family name followed by the letter A and the corresponding numeral assigned to that particular gene (Hedigar et al., 2013), e.g. *SLC22A6*. The ABC transporter genes are divided into seven families denoted the letters A-G (Vasiliou et al., 2009), followed by a number representing the gene e.g. *ABCB1*.

1.8.1 Transport of nucleotides /nucleoside

Nucleotides/nucleosides and nucleoside-derived drugs are hydrophilic molecules and diffuse, if they can, slowly across cell membranes; more often specific membrane transporters that mediate their flux across cell membranes are required (Pastor-Anglada & Perez-Torras, 2015). The genes encoding transporter proteins are highly polymorphic which therefore yields varied expression levels and enzyme-substrate specificities and thus compromises the optimal function of these proteins as well as drug efficiency (Sissung et al., 2010).

Transporter proteins have been studied lengthily *in vitro* using c-RNA injected *Xenopus laevis* oocytes and transfected mammalian cell lines and *in vivo* using animal models such as the knockout mice and many others (DeGorter et al., 2012).Transport proteins mediate the transmembrane transport of critical drugs and substances such as amino acids and nutrients by mediating their metabolism, absorption, distribution and elimination.Transport proteins controlling nucleotide entry into cells are encoded by three gene families: SLC22, SLC28, and SLC29 (Pastor-Anglada et al., 2005). Nucleotide efflux is controlled by transport proteins from the ABCB, ABCC and ABCG families (Chan, Lowes and Hirst, 2004).

1.8.1.1 SLC22 gene family

The SLC22 gene family encodes organic anion transporters (e.g., OAT1 and OAT2) and organic cation transporters (e.g., OCT1 and OCT2), implicated in the uptake of nucleoside-derived drugs (Minuesa et al., 2011). Burckhardt (2012) reported that the rat oat1 transports nucleoside prodrugs such as stavudine and zidovudine and Pastor-Anglada et al. (2005) identified the OAT1 protein (SLC22A1) as a transporter of the d4T nucleoside analogue.

1.8.1.2 SLC28 family

The SLC28 family encodes concentrative nucleotide transporters. The human concentrative nucleoside transporter protein family has three members, hCNT1, 2, and 3, encoded by SLC28A1, A2, and A3 genes, respectively. hCNT1 and hCNT2 translocate pyrimidine- and purines -, by a sodium-dependent mechanism. There is some debate whether the CNT family transport ARVs (e.g. Mangravite et al., 2003), and Minuesa et al., 2011 vs. Pastor-Anglada and Perez-Torras 2015 and Cano-Soldado et al., 2004)." hCNT1 has the capacity to transport azidothymidine and stavudine although with a low capacity (Minuesa et al., 2011). For those nucleoside analogues which are trans-located by CNT1, this occurs at a very low affinity (Cano-Soldado & Pastor-Anglada, 2012). The human CNT3 shows broad substrate selectivity and the unique ability of translocating nucleosides both in sodium- and a proton-coupled manner, and d4T is an hCNT3 substrate (Molina-Arcas et al., 2009).

1.8.1.3 SLC29 gene family

The SLC29 gene family encodes for four equilibrative nucleotide transporters (e.g., ENT1, ENT2, ENT3 and ENT4) which are involved in the mediation of facilitated diffusion and sodium independent transport (Molina-Arcas et al., 2009). The ENTs 1 and 2 are w widely expressed across human tissues (Molina-Arcas et al., 2009). The ENT3 as well ENT1 and ENT2 nucleoside transporter proteins serve as transporters of both the pyrimidines and purines with significantly varying substrate specificities (Cano-Soldado & Pastor-Anglada, 2012). ENT3 was identified as an uptake transporter of d4T (Cano-Soldado & Pastor-Anglada, 2012; Pastor-Anglada & Perez-Torras, 2015). The level of expression of the ENT3 transporter on the mitochondria is associated with the corresponding transportability of nucleoside analogue prodrugs (Minuesa et al., 2011).

1.8.1.4 ABCB family

The most studied ABC drug transporter is the p-glycoprotein (P-gp) also known as the multidrug resistance protein 1 (MDR1) efflux transporter encoded by the *MDR1* (ABCB1) gene (Michaud et al., 2012). Kis et al. (2010) identified d4T as an inhibitor of the p-glycoprotein transporter protein.

.

1.8.1.5 ABCC family: MRP5

The multi-drug resistant protein 5 (MRP5) encoded by *ABCC5* gene has been evidenced to transport d4T-MP (Marc¸al Pastor-Anglada et al., 2005). The MRP4 and 5 belong to the sub-family C of the ABC superfamily and both possess the 12 trans-membrane α-helices and are devoid of the additional membrane spanning domain at the N-terminus (Gradhand and Kim, 2008). The extent of information on the genetic variants and endogenous function of the *ABCC5* is less (Gradhand and Kim, 2008).

Fukuda and Schuetz (2012) reported that ABCC-family genes are critical in the determination of intracellular of nucleotide metabolites. Schuetz et al. (1999) first demonstrated interaction of reverse transcriptase inhibitors with MRP4 such as the monophosphorylated form of zidovudine, confirming the ability of MRP4 to interact with phosphorylated nucleosides. d4T is an MRP5 substrate, therefore it is transported out of the cell through the MRP5 efflux protein (Schuetz et al., 1999; Reid et al., 2003). The MRP5 transporter protein is expressed in the colon, liver, kidney, and skeletal muscles and consequently facilitates the efflux of d4T (Choudhuri, 2006).

1.8.1.6 ABCG family: BCRP

The (ABCG2, MXR, ABCP, BCRP) gene encodes for a breast cancer resistant protein (BCRP) a transporter protein that serves as an efflux transporter and shares the same tissue distribution with the *ABCB1* gene transporter protein (Michaud et al., 2012).

Fukuda and Schuetz (2012) reported that *ABCG2* genes are critical in the determination of intracellular nucleotide metabolites. The BCRP transporter is abundantly expressed in the liver, small intestines and lymphocytes. Cells transfected with the BCRP transporter protein showed reduced accumulation of d4T (Kis et al., 2010).

Similar transporters are used to transport nucleoside-derived drugs such as ARVs. Drug transporters can influence antiretroviral therapy outcomes in many ways: 1) bioavailability (intestinal and hepatic transporters); 2) antiretroviral penetration in sanctuary sites (e.g., brain, vaginal mucus, testicles); and 3) access in target cells (lymphocytes) (Michaud et al., 2012). **Table 1.3** shows the transporters responsible for d4T influx or efflux and which help control plasma and intracellular concentrations of d4T, as well as influencing relative levels of endogenous dN/dNTP and RN pools.

Table 1.3 The d4T nucleoside analogue drug transporters.

Transporter Protein	Influx or efflux	Reference for role of this protein in d4T transport	Abbreviation	Gene	Location
Breast cancer resistant protein	efflux	Kis et al.,2010	BCRP	*ABCG2*	4q22
Multi-drug resistant protein 5	efflux	Kis et al., 2010	MRP5	*ABCC5*	3q27.1
Concentrative nucleoside transporter 1	influx	Kis et al., 2010	CNT1	*SLC28A1*	15q25.3
Concentrative nucleoside transporter 3	influx	Minuesa et al.,2011	CNT3	*SLC28A3*	9q21.32-q21.33
Equilibrative nucleoside transporter 3	influx	Minuesa et al., 2011	ENT3	*SLC29A3*	10q22.1

1.8.2 ADME genes associated with sensory neuropathy

A number of genes involved in antiretroviral drug metabolism and transport have been implicated in the variable drug outcomes which lead to disparities in drug efficacy and the emergence of adverse drug reactions (Domingo et al., 2011; Domingo et al., 2013a). Domingo et al. (2013a) reported the association of polymorphisms in the *MTHFR* and *TS* genes with mitochondrial toxicity further mentioning that the genotypes that have a negative impact on the TS activity have been found to be associated with higher d4T-TP intracellular levels and fat redistribution. In 2011 Domingo et al. reported that *TS* genotypes associated with lessened activity of the enzyme would increase the competitive advantage of d4T-TTP against the natural dTTP pool (Domingo et al., 2011).

The TS enzyme play a key role in the denovo synthesis of pyrimidines, and the variable number tandem repeat (VNTR) polymorphism of the *TS* gene located on its promoter region has been shown to influence the d4T-TP intracellular levels (Domingo et al., 2013b). Domingo et al. (2011), reported that mutations that are associated with diminished expression of the enzyme TS are associated with increased levels of d4T-TP due to the increased competitive advantage of d4T-TP which follows upon reduced dTTP levels.

The pathogenesis of mtDNA toxicity by NRTIs involves the regulation of natural dNTP pool sizes in the mitochondria which affects mitochondrial DNA replication (Domingo et al., 2011). The elevated d4T-TP intracellular levels saturate the mitochondrial dNTP pool space and start inhibiting the mitochondrial polymerase gamma (POLG) the enzyme responsible for mtDNA replication, consequently mtDNA toxicity arises and the adverse drug reactions associated with it such as SN, lipodystrophy and are experienced.

Sterile alpha motif and histidine-aspartic domain-containingprotein 1 (SAMHD1) is a recently discovered HIV-1 host restriction factor that prevents retroviral replication at the reverse transcription stage (Ballana et al., 2014). Pauls et al. (2014) reported that dNTPs used for DNA replication are controlled by various enzymes and one of the enzymes that

tightly controls the dNTPs is the dNTP triphosphohydrolase SAM domain and HD domain-containing protein 1 (SAMHD1). SAMHD1 dephosphorylates dNTPs reducing them to concentrations below the threshold level required for RNA or DNA replication (Ballana et al., 2014).

NRTIs such as d4T compete with natural dNTPs for newly transcribed DNA at the RT step, for this reason it is hypothesized that SAMHD1 may influence HIV-1 sensitivity to NRTIs (Ballana et al., 2014). SAMHD1 targets naturally occurring dNTPs and not synthetic dNTPs such as d4T as a result this gives NRTI drugs such as d4T-TP a competitive advantage over natural dNTPs (Lewis et al., 2003). Elevated levels of intracellular d4T-TP lead to mtDNA toxicity and associated adverse events such SN.

Alterations in the metabolism of iron leads to mitochondrial dysfunction because iron plays a critical role in metabolically active cells such as neurons which require a consistent supply of iron in order to maintain events such as mitochondrial function, axonal activity and myelination (Vidal et al., 2010; Kallianpur et al., 2014). Macrophage-mediated inflammation is regulated in part by iron transport through the hepcidin pathway (Kallianpur et al., 2014).

Vidal et al. (2010) reported that heterozygous carriers of the *HFE C282Y* genotype that were receiving treatment with ddl and d4T encountered a reduced impact of sensory neuropathy in all racial groups. The *HFE H63D* polymorphism also reported a reduced occurrence of the incidence of sensory neuropathy in the univariate analysis but lost significance after adjustment for multiple testing (Kallianpur et al., 2014).

1.8.3 ADME genes associated with CD4+ T cell recovery after ART

Several genetic variants in metabolism genes of the ADME pathways have been associated with drug plasma levels and CD4+ T cell recovery after ART (Parathyras et al., 2009; Yimer et al., 2012).

Variations observed in *CYP2B6*, which encodes the predominant enzyme for the elimination of efavirenz and nevirapine ART drugs has an effect on the variable systemic exposure, plasma levels and efficacy of the treatment (Sarfo et al., 2013). However d4T is not transported by *CYP2B6*.

Rotger et al. (2006) reported that the genetic marker *CYP2B6 516T* only found within the black population was associated with high EFV levels. Elens et al. (2010) also found that the *CYP2B6 516T* polymorphism was associated with EFV plasma levels. Queiroz et al. (2017) reported that patients carrying the *CYP2B6 516T* TT genotype who use EFV, may present adverse effects such as liver or kidney toxicity that interferes with their reaction to the treatment and therefore leading to a lower CD4+ T cell count. However the patients in the current study were not using efavirenz and CYP2B6 does not metabolise d4T.

The *CYP2B6 516T* polymorphism is a missense variant that results in an amino acid change from glutamine to histidine at position 172 of the protein is associated with decreased expression and function of *CYP2B6*. Nevirapine (NVP) which is one of the highly used NNRTI drugs induces *CYP3A4* and *CYP2B6* (Arab-Alameddine et al., 2011). Saitoh et al. (2007) demonstrated that the *CYP2B6-G516T* polymorphism is associated with the pharmacokinetic of NVP and CD4 cell count response in children carrying the CYP2B6 516TT genotype.

The *MDR1/ABCB1* gene encoding for the multidrug efflux pump P-glycoprotein (P-gp) has been reported to limit the entry of protease inhibitors into CD4+ T cells (Haas, 2005). Several mutations and SNPs for the *ABCB1* gene such as 3435C>T (rs1045642) have been reported some of which affect the expression and function of the P-glycoprotein (Haufroid, 2011).

Black (Xhosa) and Coloured HIV patients of South Africa (SA) homozygous for the C allele of SNP T-129C of the *ABCB1* gene reported to have encountered elevated levels of CD4 cell count compared to patients heterozygous for the same SNP who also exhibited intermediate levels of immune recovery between that of patients having the CC

and TT genotypes (Parathyras et al., 2009). The *ABCB1* (3435C>T, exon 26) polymorphism has been associated with altered P-glycoprotein expression and is also reported to increase CD4+ T cells recovery as a result of elevated plasma concentration levels of efavirenz and nelfinavir (Haas, 2005).

1.9 AIM AND OBJECTIVES

1.9.1 Aim

To investigate the association of single nucleotide polymorphisms in d4T drug transporter genes and dNTP pool regulator genes and their association with response to d4T-ART.

The aim of the study was broken down and explained into several more specific hypotheses and aims as follows:

1. To investigate the association of selected single nucleotide polymorphisms in d4T drug transporter genes and dTTP pool regulator genes with the development of sensory neuropathy (SN) in individuals exposed to d4T, using a case-control cohort.

This work was done to investigate the hypothesis that ADME genetic variants could change the amounts of active drug available inside cells and therefore influence ART-mediated toxicity. SN is one type of ART-mediated toxicity.

2. To investigate the association of selected single nucleotide polymorphisms in d4T drug transporter genes and dTTP pool regulator genes with mtDNA copy number in blood from individuals exposed to d4T.

This work was done to investigate the hypothesis that ADME genetic variants could change the amounts of active drug available inside cells and therefore influence the amount of ART-mediated mitochondrial damage that occurs. MtDNA copy number was used as another indicator of ART-mediated toxicity.

3. To investigate the association of SNPs of d4T transport genes with cross-sectional post treatment CD4+ T cell counts in individuals exposed to d4T.

This work was done to investigate the hypothesis that ADME genetic variants could change the amounts of active drug available inside cells and therefore influence the CD4+ T cell response to ARV treatment. Only cross-sectional CD4+ T cell counts from a single time point post-ART were available for this analysis.

1.9.2 Objectives

1. To choose ~30 relevant SNPs in ADME genes to be examined in the current study.
2. To genotype chosen candidate SNPs using Sequenom assays, in samples from a case-control cohort, with and without SN after d4T use.
3. To examine statistical associations between candidate SNPs and development of sensory neuropathy.
4. To use a qPCR assay to compare mtDNA copy number in individuals with sensory neuropathy and without sensory neuropathy after d4T use.
5. To examine statistical associations between candidate SNPs and mtDNA copy number.
6. To examine statistical associations between ADME SNPs and single cross-sectional post treatment CD4 counts in this cohort.

CHAPTER 2

MATERIALS AND METHODS

CHAPTER 2: METHODS AND MATERIALS

2.1 Cohort

2.1.1 Patient recruitment

Unrelated patients of Black African descent from the Southern African Developmental Countries (SADC) were recruited at the Virology Clinic of the Charlotte Maxeke Johannesburg Academic Hospital, South Africa in 2008-2009, using the following eligibility criteria: all participants were 18 years or above, all participants were HIV positive with a confirmed diagnosis, all participants were on an antiretroviral therapy programme containing stavudine for at least six months. Blood samples were collected and DNA extraction done by Mrs A. Wadley as part of her PhD study (School of Physiology, Brain Function Research Group, University of the Witwatersrand), and the study was approved by the Human Research Ethics Committee (Medical), University of the Witwatersrand, (protocol no. M080220). Informed consent was obtained from all participants in a written format, an interpreter fluent in commonly spoken native languages and in English facilitated consent and study protocol. The study conformed to the declarations of Helsinki 2008. The number of patients initially assessed for eligibility and participation in the study were 482. A total of 404 participants were finally recruited and DNA samples from 342 patients were analysed by Wadley et al. in their studies (Wadley et al., 2011; Wadley et al., 2012; Hendry et al., 2013).

2.1.2 DNA Extraction and quantification

The initial 19 DNA samples were extracted from saliva; this was carried out using the QIAamp DNA Mini-kit (QIAGEN; Valencia, CA). The QIAamp DNA Mini-kit is a silica-membrane based DNA purification protocol that comprises of four steps carried out using QIAamp spin-columns in a micro-centrifuge tube. The rest of the DNA samples were extracted from the blood cells using the salting out method (Miller et al. 1988). This is a procedure that lyses cells using either SDS or proteinase K, followed by the salting out of proteins using sodium chloride (NaCl) from the solution and lastly the precipitation of DNA using ethanol. The DNA was later stored at -20°C.

Of the original 342 DNA samples extracted by Dr Wadley, many of the samples were sent to other collaborators or had degraded DNA. A total of 263 DNA samples that remained were used in the current study. The current study attained ethical clearance from Human Research Ethics Committee (Medical), University of the Witwatersrand, **(protocol no. M150459, Appendix A).**

The DNA quantity was determined from the concentration of the DNA sample, by measuring its absorbance at a wavelength of 260 nm using the spectrophotometer instrument (NanoDrop® ND-1000). Double distilled water was used to dilute the samples to a working concentration of 25 ng/µl. Purity was determined by calculating the ratio at absorbance 260 nm and absorbance 280 nm known as the A260/280 ratio. An absorbance ratio of 1.8-2.0 is generally accepted as pure for DNA.

2.1.3 Phenotype

Peripheral sensory data was gathered from the participants by Dr Wadley at the time of sampling using the brief peripheral neuropathy screening tool from the AIDS Clinical Trials Group (ACTG) (Chery et al., 2005).

The participants were deemed to have symptomatic HIV-SN if they experienced at least one symptom (pain, aching, burning, numbness, or pins and needles) and at least one clinical sign of neuropathy (reduced vibration sense or absent of ankle reflexes) in each leg (Wadley et al., 2011). Of the 263 participants used in the current study, 54% (n=143) were diagnosed to have symptomatic HIV-SN and therefore were assigned as cases and 46% (n = 120) reported no symptoms of HIV-SN and therefore were assigned as controls (**Figure 2.1).**

Demographic and clinical data of the patients such as age, sex, ethnicity and clinical information including duration of HIV infection, AIDS-defining illnesses, antiretroviral treatment record and other potential causes of neuropathy such as diabetes mellitus, alcoholism, vitamin B12 deficiency, exposure to isoniazid and chemotherapy were obtained. The participants' weight and height were recorded and their venous blood samples taken for hepatitis C serology.

Nadir (lowest ever) CD4+ T cell counts and CD4 + T cell counts at the time of phenotyping were recorded. Participants had spent variable lengths of time on ART (at least 6 months) and the CD4+ T cell count data were therefore cross-sectional, and not all from the same time point post- ART initiation.

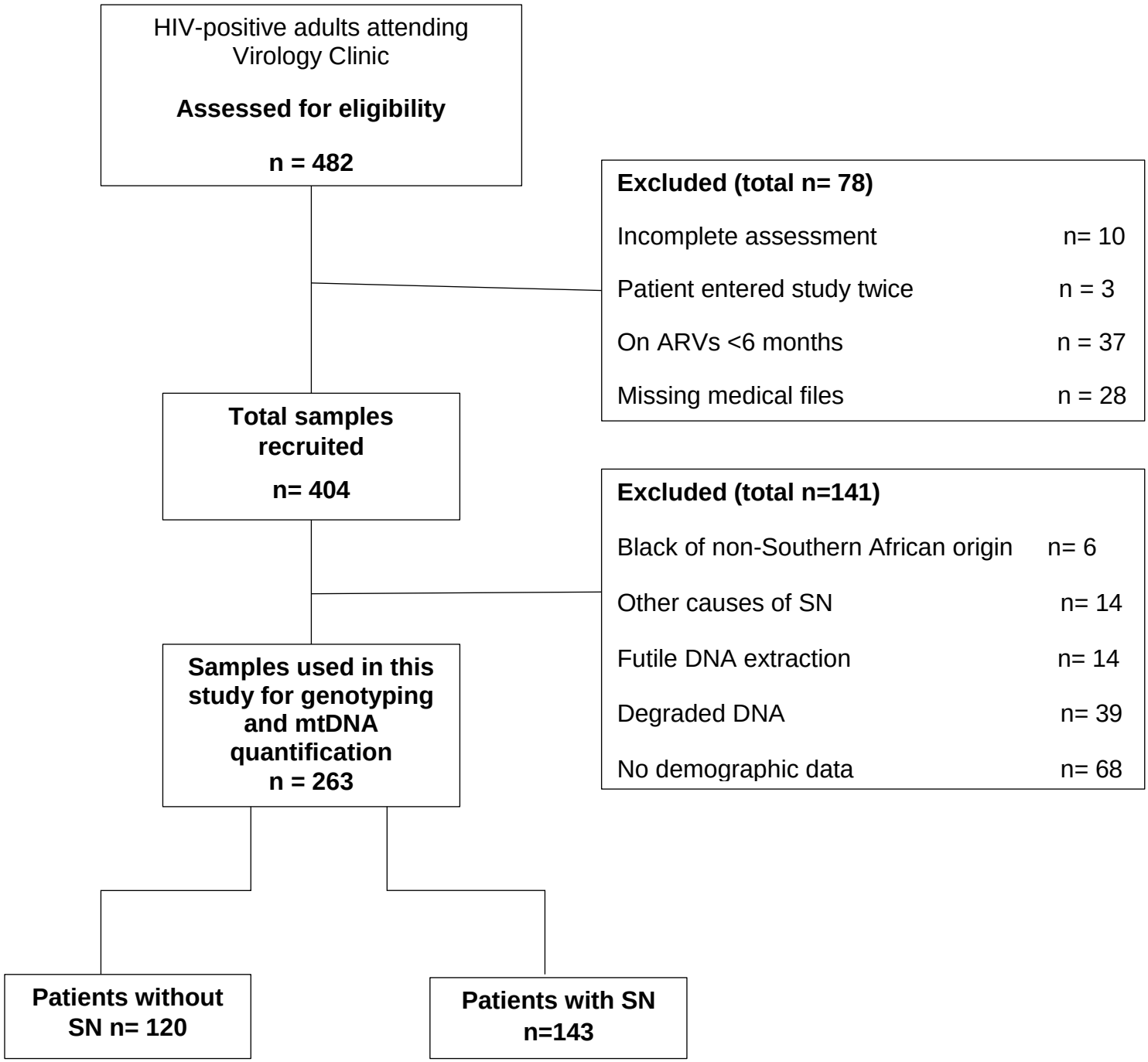

Figure 2.1 Patient inclusion and exclusion diagram, modified from Wadley et al. (2014).

2.2 SNP selection

The candidate ADME genes considered in this study were those listed in **Tables 1.2 and 1.3** in chapter 1. We chose SNPs in these ADME genes to be examined in the current study on the basis of:

- literature review of SNPs in these ADME genes
- SNP MAF≥5% in Luhya of Webuye, Kenya population
- Predicted SNP functional effect
- Use as tag SNPs
- If SNP assay could be designed and multiplexed successfully in Sequenom software.

Due to budget constraints, we genotyped 26 SNPs. Further explanation of each of these criteria is provided below.

2.2.1 Literature review of ADME genes

We used NCBI Pubmed (www.ncbi.nlm.nih.gov/pubmed), OMIM (www.omim.org), Google Scholar (scholar.google.com), Sciencedirect (www.sciencedirect.com) and Ebscohost discovery (ebscohost.com) to find literature from June 2015 to March 2016 where SNPs of interest in these genes had been previously described. SNPs of interest were those with:

- Effects on ADME gene expression
- Effects of ADME gene function
- Association with variable response to ART
- Association with variable response to other drugs
- Association with SN or any other neuropathy
- Association with mtDNA copy number

2.2.2 MAF≥5% in the LWK population

Only SNPs with minor allele frequency of at least 5 % in the Luhya of Webuye, Kenya (LWK) population were considered. This population was used as a proxy for South African populations since very little whole genome data is available for SA populations (May et al., 2013), and furthermore it has been reported by the Bioinformatics group working with the South African human genome project to be more genetically similar to the bantu populations of south Africa. The LWK allele frequency of SNPs under consideration was obtained with the use of data from the 1000 Genomes project and the International Hap Map Project. (http://browser.1000genomes.org/index.html).

The following bioinformatics tools were also used to check SNP frequencies

- (NCBI) Variation viewer (http://www.ncbi.nlm.nih.gov/variation/view/)
- Entire Genome Interface for Exploring SNPs (ENGINES) (http://spsmart.cesga.es/about.php?dataSet=engines).
- Ensembl genome browser release 87, (http://www.ensembl.org/index.html).

2.2.3 Predicted functional effect

To determine the functional significance of the SNP markers we used computational tools SIFT (http://sift.jcvi.org) and PolyPhen (http://genetics.bwh.harvard.edu/pph2/index.shtml) developed to predict the potential functional effects of the SNPs. Both these SNP prediction tools were accessed on Ensembl release 86 (http://www.ensembl.org).

2.2.4 Tag SNPs

TagSNPs selection was predicted on the linkage disequilibrium (LD) between the SNPs emanating from the same gene, and the SNP that tagged many other SNPs was prioritized. The tagSNPs were generated using the Haploview (version 4.2) software from the HapMap data set (release 24), applying the following conditions:

- Pairwise tagging
- r^2 =0.8
- Minor allele frequency (MAF) greater than 5 % in the LWK population.

2.2.5 Sequenom software

A list of SNPs was submitted to the Mass ARRAY® Assay Designer software to confirm which could be multiplexed and which could not. A single multiplex PCR reaction can contain anywhere from a few to 40 SNPs.

2.3 Genotyping using Sequenom mass-array system / -iPLEX assay

A medium-throughput SNP genotyping assay, the Sequenom Mass Array ® MALDI-TOF MS platform using iPLEX™ Gold chemistry, was used. This service was executed by Inqaba Biotec (Pretoria). Basically, the Sequenom Mass Array is somewhat similar to Single Base Extension (SBE) sequencing assays such as SnaPshot genotyping (Applied Biosystems). In the Sequenom assay, instead of single fluorescent bases being added, allele-specific mass is determined by mass spectrometry.

A total of 26 SNPs were chosen. These SNPs formed part of two multiplexes of 21 and 33 SNPs in each multiplex which also included other SNPs being examined in our research group. Mass ARRAY® System Designer software was used to design a regular pair of PCR primers for each SNP. Forward and reverse PCR primers enabled amplification of the template DNA in the region spanning the SNP region of interest. PCR fragments were designed to be short (~100 bp regions). Mass ARRAY® System Designer software was also used to design the allele specific Mass EXTEND® primers for each SNP. The extension primers **(SNPs extension primers, Appendix C)** are designed to anneal directly adjacent to each SNP position to be assayed (http://agenabio.com).

SNP genotyping with the Mass ARRAY System started with multiplex PCR amplification of short (≈100 bp) regions from template DNA. Multiplex SNP PCR was carried out in 384-well format on an Applied Biosystems Gene Amp 9700.PCR reaction and cycling conditions are shown in **Table 2.1.**

Next, shrimp alkaline phosphatase (SAP) was used to eliminate unincorporated nucleotides from the reaction. SAP protocol and temperatures are shown in **Table 2.2.**

Then a mix of different extension primers (one for each SNP position to be assayed), modified dideoxynucleotide terminators (iPLEX termination mix) and a thermo-sequenase enzyme (type of Taq) were added to the amplified target DNA. PCR protocol and cycling conditions are shown in **Table 2.3.**

During second thermocycling, the iPLEX enzyme added one adjacent terminator to each extension primer. The mass differences between each of the four modified terminators are sufficiently large to be detected by the Mass ARRAY System.

The arrays were placed into the matrix assisted laser desorption/ionisation time of flight (MALDI-TOF) mass spectrometer and the mass correlating genotypes were determined in real time. This occurs on a spectro-chip array loaded in to the analyser whereby the analyte crystals are irradiated by a laser, inducing ionisation and desorption. The time-of-flight is proportional to mass, which correlates to allele type. The mass array has been created and custom-made to detect DNA with a mass range of approximately 4500 Da to 9000 Da and can easily differentiate analytes separated by 16 Da or more. The typer software automatically generates reports that identify SNP alleles. This platform generates SNP genotypes with 99.7% accuracy.

The generated data were viewed in the Typer Software, which provided tools that allowed the evaluation and management of the results (http://agenabio.com/genetics). Starting with the known molecular mass of the input extension primers and the four terminators, the data analysis software (Typer) could determine which terminator was added to each extension primer and thus assign the nucleotide in the SNP position (Ross et al., 1998; Niels and Darnhofer-Patel, 2003; Gabriel et al., 2009; Agena Bioscience, 2015 http://agenabio.com).

Table 2.1 iPLEX multiplex PCR and reagents.

iPLEX GOLD assay PCR cycling conditions and reagents			
iPLEX-multiplex PCR			
Reagent	**Final Concentration**		**Volume**
Water, HPLC grade	N/A		1.75 µl
10 x PCR Buffer with 15 mM $MgCl_2$	1.25x (1.875 mM $MgCl_2$)		0.625 µl
25 mM $MgCl_2$	1.625 mM		0.325 µl
25 mM dNTP Mix	500 µM		0.1 µl
0.5 uM Primer Mix	0.1 µM		1 µl
5 U/µl HotStar Taq	1 Unit		0.2 µl
DNA	(5-10 ng)		1 µl
	Total volume		5 µl
Thermocycler cycling conditions	Temperature (°C)	Time	Cycle/s
	94	15 min	1
	94	20 s	45
	56	30 s	
	72	1 min	
	72	3 min	1
	4	∞	1

Table 2.2 iPLEX-SAP conditions and reagents.

SAP reaction:			
Reagent		**Volume**	
Water, HPLC grade		1.53 µl	
SAP Buff. (10 x)		0.17 µl	
SAP Enzyme (1.7 U/µL)		0.3 µl	
	Total volume	2 µl	
Thermocycler conditions	Temperature (°C)		Time
	37		40 min
	85		5 min
	4		∞

Table 2.3 Primer extension cycling conditions and reagents.

PCR reaction:				
Reagent		**Volume**		
Water, HPLC grade		0.755 µl		
iPLEX buffer plus (10x)		0.2 µl		
iPLEX terminator		0.2 µl		
Primer mix		0.804 µl		
iPLEX enzyme		0.041 µl		
	Total volume	2 µl		
Thermocycler cycling conditions		Temperature (°C)	Time	Cycle/s
		94	30 s	1
	5 X	94	5 s	40
		52	5 s	
		80	5 s	
		72	180 s	1
		4	∞	1

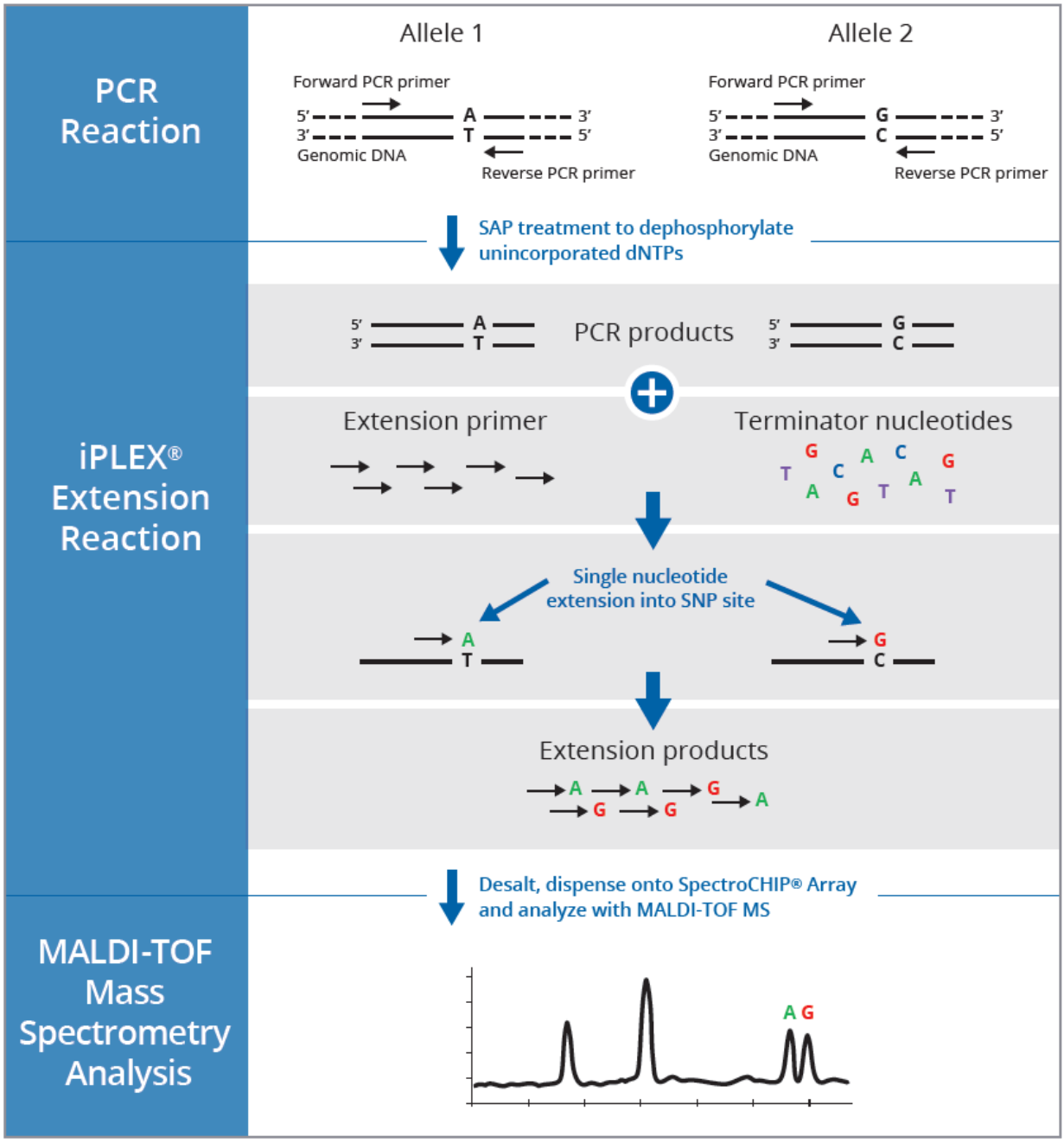

Figure 2.2 iPLEX assay (Agena Bioscience, iPLEX chemistry).

2.3.1 Quality Control of Sequenom SNP data

A standard quality control measure was performed using PLINK (Purcell et al., 2007), in order to monitor and control the quality of our genotype data. Our protocol implemented QC on a “per individual” basis and subsequently on a “per marker” basis in order to maximize the number of markers remaining in the study. The conditions applied for both the categories were the following.

i. Per individual QC
   - Identification and removal of samples with high missing genotype rates greater than 10 %.

ii. Per Marker QC
   - Identification and removal of markers with a high missing genotype rates greater than 10 %.
   - Identification and removal of SNPs showing a significant departure from Hardy-Weinberg equilibrium ($p<0.001$).
   - Removal of all SNPs with a MAF less than 1 %.

2.4 qPCR assay for mtDNA copy number quantification

Relative mtDNA copy number can be determined by comparing the amount of mtDNA present to the amount of nuclear DNA present. The mtDNA: nDNA ratio was assessed using qPCR and genomic (total) DNA extracted from whole blood.

The PCR primers for this assay were chosen using published information for the nuclear (control) gene Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) and the mitochondrial (target) *mt-ND1* gene (**see Table 2.4**). Two pairs of primers per gene were initially chosen and used for assay optimisation. *GAPDH* is known to be present at single copy in the nuclear genome (Atema et al., 2013) and no pseudogenes *of mt-ND1* are present in the nuclear genome (Grady et al., 2014).

Table 2.4 PCR primers used for optimization of mtDNA copy number assay.

Primer name	Sequence (5´-3´)	Length	Tm, °C	Direction	Amplicon size	Reference
***GAPDH*1**	AAATCAAGTGGGGCGATGCTG	21	61.56	Forward	208	**Zhang et al., 2005**
	GCAGAGATGATGACCCTTTTG	21	56.95	Reverse		
***GAPDH*2**	TGACAACGAATTTGGCTACA	20	55.60	Forward	197	**Cooray et al., 2002**
	GGGGTCTACATGGCAACT	18	56.17	Reverse		
***mt-NAD*2A**	CTAGCCCCCATCTCAATCATA	24	63.24	Forward		**Bonod-Bidaud et al., 2001**
	GAATGCGGTAGTAGTTAGGAT	24	64.45	Reverse	243	
***mt-NAD*2B**	ATGGCCAACCTCCTACTCCTCATT	24	63.24	Forward	154	**Psarra and Sekeris, 2011**
	TATGGCGTCAGCGAAGGGTTGTA	23	64.11	Reverse		

2.4.1 Optimization of the mtDNA copy number quantification assay

2.4.1.1 Endpoint PCR

The qPCR mtDNA copy number primers in Table 2.4 were optimised using a gradient PCR, in order to assess specificity, determine the ideal PCR master mix protocol and decide on the optimal annealing temperature. This optimisation was initially performed using regular (endpoint) PCR. The DNA was amplified using KAPA Taq ready mix PCR kit (Kapa Biosystems) with a total reaction volume of 20 µl. The optimized reaction master mix contained the following components **(Table 2.5):**

Table 2.5 PCR reaction mix

Component	25 µl reaction	Final concentration
2X KAPA Taq Master Mix	12.5 µl	1X
10 µM Forward primer	0.6 µl	0.24 µM
10 µM Reverse primer	0.6 µl	0.24 µM
Distilled water	10.3 µl	N/A
Template DNA	1 µl	0.8 ng/µl

N\A –Not applicable

A gradient PCR using a Biorad gradient thermal cycler was used for primer optimization. The cycling protocol for all the primers were the same except for the annealing temperatures. The cycling conditions consisted of an initial denaturation step at 95 °C for 10 min followed by 30 cycles of denaturation at 95 °C for 1 min, annealing at 45 °C-50 °C, 48 °C-54 °C for 30 s and extension at 72 °C for 1 min with a final extension step for 2 min.

Agarose gel electrophoresis was used to visualize and size the PCR products and confirm PCR specificity. 1 % Agarose gel, prepared by weighing 0.5 g of agarose powder (SeaKem® LE Agarose) and dissolving it in 50 ml of 1X Tris-Borate-EDTA (TBE) buffer (Life Technologies or Ambion™) was used to separate the PCR products by electrophoresis. The agarose powder was dissolved by microwaving in the 1X TBE buffer for 2-3 minutes, subsequently the liquid gel was cooled and stained with 5 µl GR green (Biotium) and cast on a tray to set for 15 to 20 minutes. 5 µl of the PCR products were loaded on to the gel by mixing it with the loading dye (Thermofisher Scientific). A DNA ladder (Kapa Biosystems) was used to scale the size of the bands of the PCR products. The agarose gels were electrophoresed for 30 minutes to an hour in 1X TBE at 100V.The agarose gels were visualized and images taken by Gel doc (Biorad Chemidoc MP system) using the Image Lab software.

2.4.2 Optimization of qPCR

PCR quantification for *GAPDH* and *mt-ND1* genes was performed using the Light Cycler 480 instrument II (Roche Diagnostics, Indianapolis, USA) according to the manufacturer's instructions.

2.4.2.1 Principle of the SYBR green assay

SYBR green I is an inexpensive non-sequence specific double stranded DNA (dsDNA) binding fluorescent dye. SYBR green I binds randomly to any dsDNA. When SYBR green I is free in solution it barely fluoresces but when intercalated to a dsDNA its fluorescence emission increases to over a 100-fold relative to the quantity of the dsDNA present in the solution. After denaturation all the DNA becomes single stranded **(Image a, Figure 2.3)**, SYBR is unbound and the fluorescent intensity is low. During the annealing stage the PCR primers hybridize to the target sequence creating small regions of dsDNA that SYBR green I can bind, and thereby elevating the fluorescence intensity **(Image b, Figure 2.3).**

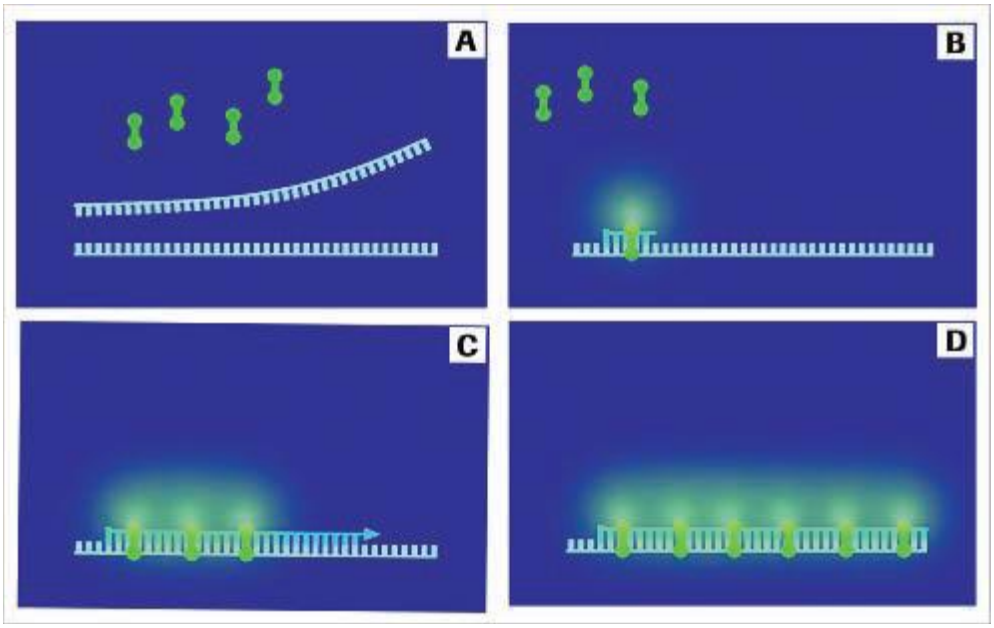

Figure 2.3 qPCR protocol using SYBR green I (Image A-D) (Roche applied science, 2004).

During the extension step of PCR, the primers are extended which allows more SYBR green I to bind **(Image c, Figure 2.3).** At end of the extension/elongation phase the entire DNA becomes dsDNA which increase the amount of SYBR green I that is bound **(Image d, Figure 2.3).**

2.4.2.2 mtDNA: nDNA assay

GAPDH and *mt-ND1* were quantified separately, each in a 25 µl reaction volume with 0.2 µM for the reverse and forward primer, 12.5 µl Maxima (2X) SYBR green I qPCR master mix (Thermo Scientific), 1 µl of 25 ng/µl DNA concentration and 10.3 µl of double distilled water. Amplification was carried out using a three-step-cycling protocol according to the specifications of the Maxima SYBR green qPCR master mix. *GAPDH* and *mt-ND1* were amplified separately and two technical repeats were done for each gene.

Table 2.6 qPCR conditions for mtDNA: nDNA quantification.

Step	Temperature, ºC	Time	Number of Cycles
Initial denaturation	95	10 min	1
Denaturation	95	15 s	40
Annealing	50 –*GAPDH* 58 –*mt-ND1*	30 s	
Extension	72	30 s	

Data acquisition was performed during the extension step.

2.4.2.3 Melting curve analysis

PCR was followed by a melting curve analysis to determine the specificity of the PCR reaction. The melting was performed after the amplification cycles and continued for one cycle at three melting temperature targets, which were: 97 ºC for 1 minute at a ramp rate of 4.4, 40 ºC for 10 seconds at a ramp rate 2.2 and 95 ºC acquisition mode with 5 acquisitions per ºC at a ramp rate of 1.1.

2.4.2.4 Standard curves

Standard curves were used to determine the PCR efficiencies of the primers. A series of tenfold-dilutions was prepared using the genomic DNA from HEK293 cells starting with the initial concentration of 20 ng/µl yielding the following DNA dilutions 2 ng/µl, 0.2 ng/µl and 0.02 ng/µl. PCR efficiencies determine the amount of target DNA that is amplified per cycle of PCR (Svec et al., 2015).

2.4.2.5 Crossing Points

Crossing point (Cp) also referred to as the threshold cycle (Ct) is defined as the fractional cycle number or crossing point whereby the reporter dye exceeds or crosses the threshold point. The Cps were determined by using the second derivate maximum method, it is a point that correlates with the maximum of the second derivative of the amplification. The threshold point was determined automatically by the Roche Light Cycler system.

2.4.3 Assay validation

We validated our qPCR assay using cell lines depleted of mtDNA. It is known that treatment of eukaryotic cells with ethidium bromide (EtBr) depletes endogenous mtDNA while nDNA is maintained (creating so-called ρ0 cells devoid of mtDNA) (Seidel-Rogol and Shadel, 2002). We therefore treated Human embryonic kidney (HEK) 293 cells with EtBr for 6 days to ensure that our assay could detect depletion of mtDNA content.

HEK293 cells were grown in Dulbecco's modified Eagle's medium (DMEM) high glucose (4.5 g/l) (Invitrogen Gibco) supplemented with 10 % fetal bovine serum, non-essential amino-acids and penicillin/streptomycin at 37 °C in an atmosphere humidified with 5 % CO_2.The cells were grown to a confluence of 50-80 % prior treatment with Ethidium bromide (EtBr).Treatment of the cells with EtBr to hinder mtDNA replication, the cells were grown in the same media as described above with the presence of 50 ng/ml EtBr. The cells were grown for six days and harvested at three different times. The first time point was prior treatment (Day 0), the second time point was 48 hours post treatment (day 3) and the last day of harvest was 96 hours post treatment (day 6). The cells harvested were stored in the -20 °C freezer before DNA extractions were performed.

DNA extraction from the HEK293 cells, from the three time-points was done using the DNA extraction kit (Life Technologies), following the manufacturer's instructions, starting from step 8 of the protocol; protease in nuclear lysis buffer was added to the pellet of HEK293 cells followed by vortexing and heating at 65 ºC for 10 min in order to lyse the cell membrane and breakdown the proteins.

The protein clearing solution was added to denature and remove proteins bound to the DNA; this was followed by incubation at 65 ºC on a heating block for 10 minutes in order to fully residue the crude lysate. This was followed by an incubation step at 0-5 ºC, on ice and centrifugation of the tube for 5 minutes at > 12,000 × g, with the DNA contained in the supernatant transferred to a different tube.

To precipitate the DNA, ice cold (0–5 °C) 90 % ethanol was added to the supernatant containing the DNA and the tube was inverted about 10 times and a stringy, white to translucent mass of DNA appeared. The DNA pellet was centrifuged at >12,000 × g for 5 minutes followed by the removal of the supernatant, the DNA pellet was then washed using 70 % (v/v) ethanol and centrifuged for 1.5 minutes at >12,000 × *g*. The ethanol was then decanted and the DNA pellet was set to air dry by inverting the tube with the cap open over a clean lint-free disposable laboratory wipe (Kim Wipe®) for 2–3 minutes to evaporate the ethanol. A cotton swab was used to adsorb the residual ethanol from the sides and top of the tube. DNA was dissolved in sterile distilled water. To fully dissolve the DNA, the tube containing the DNA was placed on a heating block for 10 minutes at 65 ºC. DNA was stored in the -20 ºC freezer.

The DNA was quantified on the qPCR machine (Roche Diagnostics, Indianapolis, USA) using the method described above **(2.4.2.2 and 2.4.2.5).**

2.4.4 qPCR Plate designs for cohort analysis

The relative quantification of the mtDNA copy number for the 263 cohort samples was carried out in a total of 14 barcoded white 96-well plates (Biorad California 94547, USA) using optical seals for covering. Seven plates were used for the *mt-ND1* gene and the other seven for the *GAPDH* gene. The arrangement for plates was as follows.

- A series of dilutions for standard curves, this occurred in a duplicate fashion as two technical repeats were run for each sample to account for errors and ensure reproducibility of the results.
- SN samples and/or the control samples in duplicate

For each sample, Ct values were obtained for each of the genes and were averaged across replicate wells. The fold change in mtDNA: nDNA ratio in treated vs untreated samples was calculated as $2^{-\Delta\Delta Ct}$ where:

$\Delta Ct = [Ct\ (target) - Ct\ (ref)]$

Where:

Ct (target) = Ct value of gene of interest (*mt-ND1*) in untreated sample

Ct (ref) = Ct value of control gene (*GAPDH*) in untreated sample.

The following QC criteria were applied to the samples during data analysis:

- All samples that had no amplification, either due to low or no DNA in respective wells were excluded for analysis.
- Replicate Samples that had a Ct difference of more than 2 were excluded from analysis.
- Samples with average Ct of greater than 30 for both genes (*GAPDH* and *mt-ND1*) were excluded for analysis.

2.5 Statistical analysis

2.5.1 Demographic data and SN

Statistical analyses to determine the effect of the demographic and clinical variates on individuals with SN (cases) and individuals (without SN) as controls were done using GraphPad Prism 7 software (Graphpad.com). For continuous variables a non-parametric Mann-Whitney test and linear regression analysis were used and the chi-squared test was used for categorical variables. The p-value ($p<0.05$) was considered to be significant for all the statistical tests performed in this thesis.

2.5.2 mtDNA copy number analysis

A non-parametric Mann-Whitney test using GraphPad Prism was used to analyse the statistical differences between the mtDNA/nDNA ratios in SN cases vs. controls and significance was regarded as ($p<0.05$).

2.5.3 SNP analysis

The statistical analysis for SNPs, the construction of haplotype blocks and linkage disequilibrium analysis were performed using the open source software PLINK version 1.07, with the support of the Haploview software (Purcell et al., 2007).

PLINK is a freely available, open source whole genome association toolset widely used for large scale genetic association analysis. PLINK is primarily designed for analysing genotype/phenotype data in case-control association tests. The five domain functions of PLINK are data management, summary statistics, assessment of population stratification, association analysis and Identity by descent (IBD) estimations. PLINK is a command line program written in C/C++, whereby all commands involve typing at the command prompt (Purcell et al., 2007).

2.5.3.1 Allele, genotype and haplotype frequencies in the SA cohort

After QC, allele and genotype frequencies in the entire cohort were calculated. The allele frequencies in this South African cohort were compared to those in other African populations, i.e. the Luhya of Webuye, Kenyan (LWK) and the Yoruba in Ibadan, Nigerian populations using data from 1000 Genome phase 3 data accessible on the Ensembl database (Ensembl release 87, http://www.ensembl.org/index.html, January 2017).

More than one SNP per gene was studied for the following seven genes:

- *ABCG2* (4 SNPs)
- *SLC28A1*(3 SNPs)
- *SLC28A3* (2 SNPs)
- *MTHFR* (2 SNPs)
- *RRM1* (4 SNPs)
- *TK2* (3 SNPs)
- *SAMHD1* (3 SNPs)

Haplotypes and LD across SNPs of these genes were examined using PLINK version 1.07 (Purcell et al., 2007) and Haploview version 4.2 (Barret et al., 2005). Pairwise linkage disequilibrium of genes with more than one SNP genotyped was calculated. SNPs were considered to be in LD if $R^2 \geq 0.2$ and were less than 500 Kb apart. The files PLINK.info and PLINK.ped generated through PLINK were used on the Haploview software to create a graphical representation of the LD pairwise calculations. Haplotypes for the SNPs were constructed and their frequencies estimated using the sliding window specification approach in PLINK.

2.5.3.2 Association analysis with phenotypes

PLINK was used to look for associations between candidate SNPs and phenotypes, and the phenotypes considered included:

- **SN ever** – sensory neuropathy ever was defined as having at least one sign and at least one symptom of sensory neuropathy as outlined in the brief peripheral neuropathy screening tool from the AIDS Clinical Trials Group (ACTG) (Chery et al., 2005), during the time of screening or at previous times.

- **mtDNA copy number**- mtDNA copy number phenotype was taken as a ratio between mitochondrial gene copies cross points and the nuclear gene copies cross points from qPCR assay (mtDNA: nDNA).

- **CD4+ T-cell count (current)** – CD4+ T-cell count was the number of CD4+ T-cells per microliter (cells/µl) of the blood sample at the time of sample collection.

The SNP association analyses were carried out using both univariate and multivariate analysis. All 26 SNPs were tested for association with the phenotype in both univariate and multivariate analysis.

2.5.3.2.1 Univariate analysis

The association of individual SNPs with phenotype data such as SN-ever, or mtDNA copy number or CD4+ T-cell count were analysed using the univariate data analysis method. The empirical tests that we have used in this study are the following.

- Allelic test
- Genotypic (2 df) test
- Dominant gene action (1df) test
- Recessive gene action (1df) test
- Haplotype association test

Three different genotypic models of inheritance were assessed as follows: The dominant model assessed having at least one minor allele versus not having any minor allele, the recessive model assessed having two minor alleles versus having at least one major allele, and the codominant model assessed all genotypes separately **(Table 2.7).**

Empirical p-values (pEMP) were generated by PLINK, using a Monte Carlo-based max (T) permutation testing method. A hundred thousand simulations were performed. PLINK generated both uncorrected (EMP1) and corrected (EMP2) empirical p-values. The corrected empirical p-value (EMP2) does not assume that the tests are independent, but it still controls the probability of observing at least one false positive per experiment and hence is more stringent than the conventional EMP1 p-value. In both cases, p-values of <0.05 were taken as significant (Hider et al., 2008).

Table 2.7 Allelic and Genotypic models used in association tests.

Test			
Allelic:	D	versus	d
Dominant:	(DD, Dd)	versus	dd
Recessive:	DD	versus	(Dd, dd)
Genotypic: or codominant	DD	versus	Dd versus dd

‘’D’’- Minor allele and ‘’d’’ Major allele

2.5.3.2.2 Multivariate analysis

Two demographic factors which were significantly associated with SN (age and height) were used as covariates in multivariate analyses of SNP association with sensory neuropathy.

One clinical variable which was significantly associated with CD4+ T cell count (number of months HIV+) was used as a covariate in the multivariate analysis of SNP association with CD4+ T cell count.

No demographic or clinical covariates were identified as associating with mtDNA copy number, therefore multivariate analysis of association between genetic variation and mtDNA copy was not performed.

A SNP-based logistic regression was used for the association analysis with sensory neuropathy ever (SN), as it is presented as a dichotomous phenotype in our study. Linear regression models for continuous traits (quantitative) were used for SNP associations with phenotypes CD4+ T cell count and mtDNA copy number.

CHAPTER 3

RESULTS

CHAPTER 3: RESULTS

3.1 Demographic and clinical data and their association with sensory neuropathy and CD4+ T cell count

Two hundred and sixty-three (n= 263) Black adult participants of African descent were enrolled in this study, 205 (78 %) were female and 58 (22 %) were male. The participants were all from southern African countries with over 95 % of the participants being South Africans from all the 9 indigenous ethnic groups in South Africa (Venda, Zulu, Pedi, Sotho, Tswana, Swati, Tsonga, Xhosa, and Ndebele) with the other remaining participants from the Niger-Kordofanian-speaking countries (Zimbabwe, Mozambique, Malawi).

The demographic and clinical differences between those with sensory neuropathy (SN) and those without SN were analysed. The prevalence of SN was not associated with the following clinical variables: CD4+ T cell count at time of sample collection; CD4+ T cell count nadir, number of months HIV+ and demographic variables; gender, weight and BMI. Height and age were significantly associated with sensory neuropathy during the univariate evaluation between SN and SN-free individuals, with $p<0.0001$ and $p<0.007$ respectively, as per Table 3.1. The mean (range) height for all the participants was 158 cm (132-186.6) with the cases (with SN) having a mean of 159.4 cm and the mean for the control group (without SN) 156.4 cm. The mean age of the participants was 40.5 (20-58) years for the cases and 35.86 (19-59) years for the controls.

Descriptive statistics for the cohorts used in this study are provided in **Table 3.1,** with females representing 77 % and 79 % of the SN cases and controls respectively. The CD4+ T cell count was generally high post-ART with a mean (range) of 429.2 (0-1091) cells per mm^3.

Table 3.1 Associations between demographic and clinical factors with SN.

Demographic description	Cases with SN	Controls without SN	Total	p-value
Sample Size	143	120	263	-
Female	110 (77 %)	95 (79 %)	205 (78 %)	0.333
Age in years Mean	40.5	35.86	38.38	**<0.0001**
Height in cm Mean	159.4	156.4	158.1	**0.007**
Weight in kg Mean	69.56	67.49	68.61	0.231
Body mass index (BMI) Mean	27.45	27.66	27.55	0.970
Months since HIV diagnosis Mean	53.58	49.32	51.63	0.538
CD4 count (current) Mean	432.5	425.3	428.9	0.392
CD4 count (nadir) Mean	86.7	94.16	90.1	0.343

(Red) p<0.05 significantly associated with sensory neuropathy.

-- Not applicable/No data.

All the other p-value tests were obtained using the t-test non-parametric Mann Whitney test.

None of the demographic data was significantly associated with mtDNA copy number in our study cohort as shown by **Table 3.2**.

Therefore no covariates were identified that needed to be taken into account in multivariate analyses of genetic variation associated with mtDNA copy number.

Table 3.2 Association of demographic data with mtDNA copy number.

Demographic variable	p-value
Age	0.470
Height (cm)	0.947
CD4 Nadir (cells.mm^3)	0.409
Gender (%)	0.251
CD4 at Sampling (cells.mm^3)	0.515
Months HIV-positive	0.414
Months on d4T	0.773

After performing linear regression analyses between CD4+ T cell count and clinical and demographic variables only the number of months HIV+ emerged as an independent variable associated with CD4+ T cell count ($p<0.0001$).

The demographic variables not associated with CD4+ T cell count were age, height, weight, BMI and the clinical variable CD4+ T cell count (nadir).

Table 3.3 Associations between demographic and clinical variables with CD4+T cell counts at time of sample collection.

Demographic description	n/mean	p-value	R-square
Size (n)	263	-	-
Female (n)	205	-	-
Age (years)	38.38	0.097	0.011
Height (cm)	158.1	0.911	0.00005
BMI (kg/m^2) mean	27.55	0.286	0.004
Weight (kg) mean	68.61	0.2704	0.005
CD4+ count nadir (cells\mm^3)	90.1	0.264	0.005
Months HIV+ mean	51.63	**<0.0001**	0.102

-- Not applicable/No data.

(Bold) $p<0.05$ significantly associated with CD4 count restitution.

3.2 Cohort DNA Quantification

The cohort DNA samples had high yield DNA concentrations with the lowest having the concentration of 56.59 ng/µl and the highest being 2395.20 ng/µl with the average concentration for all the cohort samples at 494, 33 ng/µl. Good quality DNA has A260/280 ratio between 1.7-2.0, the cohort samples had A260/280 ratio averaging 1.70 which is highly acceptable and regarded as pure and suitable for further downstream applications.

3.3 SNP Choice

We chose SNPs in candidate ADME genes to be examined in the current study on the basis of the criteria below as described in the Methods section:

- literature review of SNPs
- MAF≥ 5 % in Luhya of Webuye, Kenya population in 1000 genomes and HapMap data
- Predicted functional effect
- Use as tag SNPs
- If the SNP assay could be designed and multiplexed successfully in the Mass ARRAY designer PCR software.

The SNPs were prioritized based on meeting the majority of the criteria, with the ability of the SNP assay and multiplex as the cut-off point.

Results showing many interesting SNPs meeting these criteria are summarised in the **Table 3.4** below. From this list, a total of 26 SNPs were finally chosen for genotyping in the Sequenom Mass ARRAY iPLEX genotyping platform.

Table 3.4 SNPs of interest meeting our criteria for SNP selection, with 26 indicated by * selected for genotyping.

Gene	SNP	Predicted effect on function.		Function	Reference	MAF in LWK	Tag SNP
		SIFT	PolyPhen				
SLC28A1	rs2290272*	Tolerated	Benign	Reduced expression and lowered affinity for substrates.	Kroetz, Yee and Giacomini, 2010	0.172	Tag (3)
SLC28A1	rs2242046*	Tolerated	Benign	A synonymous variant resulting in amino-acid change from aspartate to asparagine that is associated with dose-toxicity of zidovudine.	Tekola-Ayele et al., 2015	0.06	Tag (1)
SLC28A1	rs8187758*	Tolerated	Possibly damaging	A missense variant, increases thymidine uptake.	-	0.131	Not tag
SLC28A3	rs4877847*	-	-	An intron variant; loss of function for the influx transporter.	Visscher et al., 2011	0.438	Tag (3)
SLC28A3	rs7853758*	-	-	A synonymous variant; predictive of anthracycline-induced cardiotoxicity and reduced *SLC28A3* mRNA expression.	Visscher et al., 2011	0.32	Tag (6)
ABCC5	rs3749442*	-	-	Associated with vincristine-induced peripheral neuropathy.	Broyl et al., 2010	0.237	Tag (45)
ABCG2	rs2725252*	-	-	An intron variant; affects mRNA expression of *ABCG2*.	Delord et al., 2013	0.206	Not tag
ABCG2	rs12505410*	-	-	An intron variant; affects mRNA expression of *ABCG2*.	Delord et al., 2013	0.126	Not tag
ABCG2	rs3114018*	-	-	Associated with severe oxaliplatin-induced peripheral neuropathy.	Custodio et al., 2014	0.333	Not tag
ABCG2	rs2622604*	-	-	Reduced *ABCG2* (BCRP) mRNA expression.	Cha et al.,2009	0.126	Tag (7)

Table 3.4 continued. SNPs of interest meeting our criteria for SNP selection, with 26 indicated by * selected for genotyping.

Gene	SNP	Predicted effect on function		Function	Reference	MAF in LWK	Tag SNP
		SIFT	Polyphen				
RRM1	rs12806698*	-	-	Associated with expression levels.	Cao et al.,2013	0.015	Tag(1)
RRM1	rs1465952*	-	-	Associated with expression levels.	Cao et al.,2013	0.361	Tag (18)
RRM1	rs11030918*	-	-	Associated with expression levels.	Cao et al.,2013	0.189	Not tag
RRM1	rs1042927*	-	-	Associated with expression levels.	Cao et al.,2013	0.232	Tag (73)
RRM1	rs1561786[nt]	-	-	Promoter SNP associated with lower levels of RRM1 gene expression.	Cao et al.,2013	0.026	Tag (3)
RRM2B	rs16918482*	-	-	3'UTR variant.	-	0.076	Tag (25)
MTHFR	rs1801131* .	Tolerated	Benign	Decreased activity of *MTHFR*; associated with SN and pancreatitis.	Domingo et al., 2013a	0.175	Tag (2)
MTHFR	rs1801133*	Deleterious	Possibly damaging	Decreased activity of *MTHFR*; associated with SN and pancreatitis.	Domingo et al., 2013a	0.133	Not tag
TK2	rs3743712*	-	-	Affects *TK2* expression, 3'-UTR variant.	Gamazon et al., 2012	0.369	Tag (15)
TK2	rs11859474*	-	-	An intron variant.	-	0.293	Tag (33)
TK2	rs2288399*	-	-	An intron variant.	-	0.08	Tag (23)
DHFR	rs1650723*	-	-	An upstream gene variant.	-	0.071	Tag (7)
SLCA19	rs1051266*	Tolerated	Benign	A missense variant, associated response to methotrexate.	de Rotte et al., 2012	0.308	Tag (6)

Table 3.4 continued. SNPs of interest meeting our criteria for SNP selection, with 26 indicated by * selected for genotyping.

Gene	SNP	Predicted effect on function		Function	Reference	MAF in LWK	Tag SNP
		SIFT	PolyPhen				
SAMHD1	rs1291142*	-	-	An intron variant, affects mRNA expression.	Coon and Wang, 2012	0.237	Tag (19)
SAMHD1	rs1891643*	-	-	An intron variant.	-	0.207	Tag (12)
SAMHD1	rs8124728*	-	-	3'UTR variant.	-	0.152	Tag (3)
RRM2	rs7574663*	-	-	Promoter SNP associated with levels of *RRM2* gene expression.	Cao et al.,2013	0.16	Tag (4)
ABCC5	rs3749445[nt]	-	-	oxaplatin response , 3'UTR	-	0.121	Tag (2)
ABCC5	rs562[nt]	-	-	3' UTR; irinotecan-toxicity.	-	0.343	Not tag
ABCC5	rs939336[nt]	-	-	-	-	0.278	Tag (32)
ABCG2	rs2231137[nt]	Tolerated	Benign	Reduced level of protein, decreased enzyme activity.	Hampras et al., 2010	0.116	Tag (1)
ABCG2	rs7699188[nt]	-	-	Predicted to result in the gain of an HNF4 site.	Rudin et al.,2016	0.407	Tag (24)
SLC28A3	rs56350726[nt]	Deleterious	Possibly damaging	Associated with gemcitabine toxicity.	-	0.07	Tag (5)
SLC28A3	rs10868138[nt]	Deleterious	Benign	Associated with gemcitabine toxicity.	-	0.17	Tag (3)

*genotyped

nt = not genotyped in this study

"-"score not applicable

MAF = minor allele frequency

LWK = Luhya of Webuye, Kenya

3.4 Sequenom Results

Figure 3.1 shows an example of a Mass Array spectrum multiplexing some of the candidate SNPs chosen for genotyping in this study. From this Mass Array Spectrum, using the Typer Software a report was generated which was able to distinguish between SNP alleles within each sample interrogated. Here SNP rs9905016 from *POLG 2* was used as an example. **Figure 3.2** shows an example of an rs9905016 heterozygous sample with two peaks at 5330 and 5410 Da respectively while **Figure 3.3** shows an example of an rs9905016 homozygous sample with only one peak at 5330 Da.

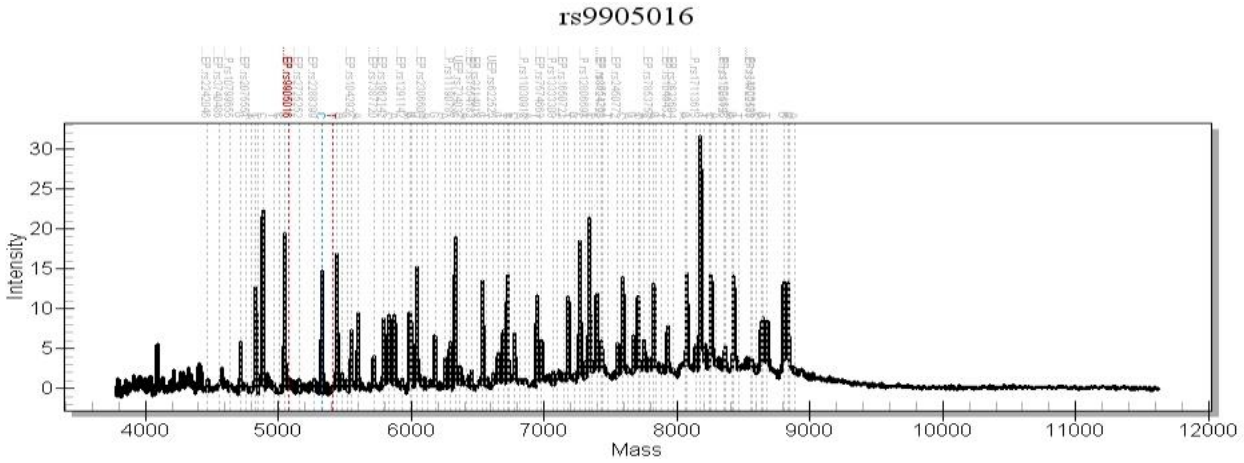

Figure 3.1 Mass spectrums for all the SNPs in the multiplex reaction.

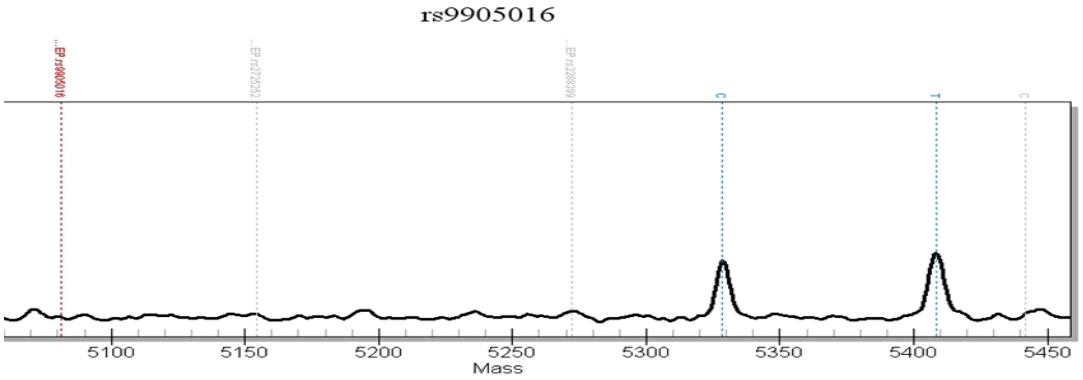

Figure 3.2 Example of mass spectrum for rs9905016 Heterozygote, with two peaks at 5330 and 5410 Da respectively.

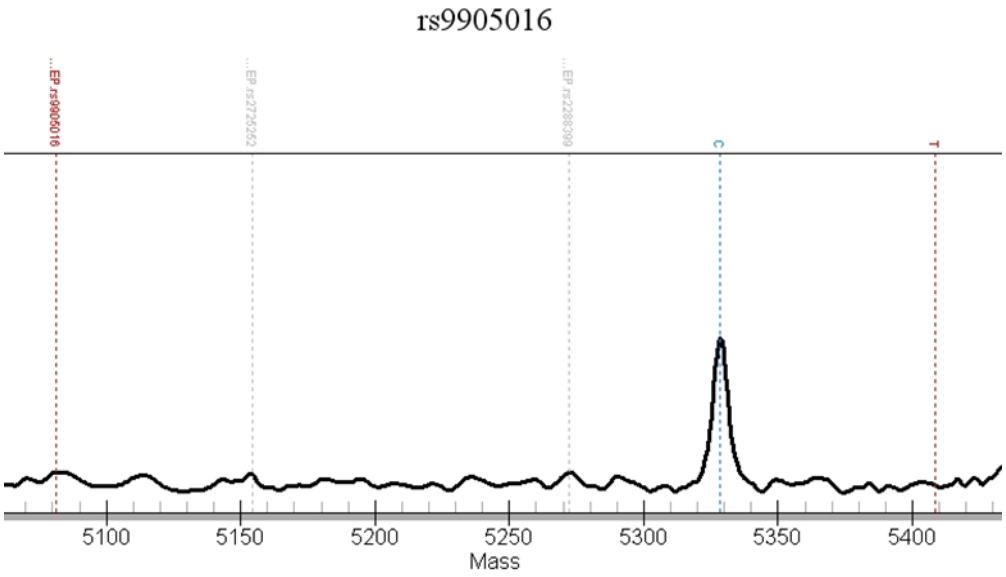

Figure 3.3 Example of mass spectrum for rs9905016 homozygote with only one peak at 5330 Da.

3.4.1 Minor allele frequencies in the SA population

After QC 243 samples remained (120 controls and 143 cases) with a genotyping rate of 99%. The minor allele frequencies (MAFs) in the cohort were calculated using the PLINK toolset, for the 26 SNPs genotyped and all of the SNPs had a frequency greater than 1 % **(See Table 3.5).** All the 26 SNPs were in Hardy –Weinberg equilibrium after comparing the genotype expected and observed p-values in the cases and controls. All the SNPs passed the genotype missingness test set to exclude all SNPs missing 10 % of the genotypes. The *RRM1* gene SNP rs12806698 allele (A>C) had the lowest frequency 0.021 in the current study.

The *MTHFR* SNPs allele frequencies as discovered by Scholtz et al. (2002) in a black population that comprised of three SA ethnic groups of Zulu, Pedi and Xhosa differed significantly to the allele frequencies that we reported in this study, at 0.04 vs. 0.07 and 0.09 vs. 0.144 for SNP rs1801133 and rs1801131 respectively.

Table 3.5 Minor allele frequencies (MAFs) for d4T transporter genes and dNTP regulator genes in the SA cohort.

Chr[a]	Gene	SNP ID	Position[b]	A1[c]	A2[c]	SA MAF[d]
d4T transporter genes						
3	*ABCC5*	rs3749442	183942797	A	G	0.208
4	*ABCG2*	rs2725252	88140758	A	C	0.193
4	*ABCG2*	rs2622604	88157772	T	C	0.132
4	*ABCG2*	rs12505410	88109689	G	T	0.122
4	*ABCG2*	rs3114018	88143429	C	A	0.419
9	*SLC28A3*	rs4877847	84331502	C	A	0.469
9	*SLC28A3*	rs7853758	84286011	A	G	0.5
15	*SLC28A1*	rs2242046	84935498	A	G	0.025
15	*SLC28A1*	rs2290272	84904200	A	G	0.242
15	*SLC28A1*	rs8187758	84905644	A	C	0.155
dNTP pool regulator genes						
1	*MTHFR*	rs1801131	11794419	G	T	0.144
1	*MTHFR*	rs1801133	11796321	A	G	0.070
2	*RRM2*	rs7574663	10119200	G	C	0.084
5	*DHFR*	rs1650723	80626211	T	C	0.062
8	*RRM2B*	rs16918482	102206460	C	A	0.064
11	*RRM1*	rs1465952	4093309	A	G	0.477
11	*RRM1*	rs11030918	4094257	C	T	0.200
11	*RRM1*	rs12806698	4094744	A	C	0.021
11	*RRM1*	rs1042927	4138699	C	A	0.181

Table 3.5 continued. Minor allele frequencies (MAFs) for d4T transporter genes and dNTP regulator genes in the SA cohort.

Chr[a]	Gene	SNP ID	Position[b]	A1[c]	A2[c]	SA MAF[d]
16	*TK2*	rs3743712	66509747	G	A	0.391
16	*TK2*	rs11859474	66520637	A	G	0.263
16	*TK2*	rs2288399	66531589	G	A	0.152
20	*SAMHD1*	rs8124728	36892303	A	G	0.169
20	*SAMHD1*	rs1291142	36896959	G	A	0.231
20	*SAMHD1*	rs1891643	36941608	G	A	0.187
21	*SLC19A1*	rs1051266	45537880	C	T	0.311

[a] Chromosome number

[b] Position of the SNP on the corresponding chromosome number as prescribed by the Ensembl database.

[c] Alleles assigned by 1000 genomes phase 3 database.

[d]The MAF from the Southern African population data (cohort samples genotyped in this study).

3.4.2 Minor allele frequencies in SA vs other African populations

The minor allele frequencies for the Southern African cohort for 26 SNPs were compared with the MAF in other African populations, the LWK population and the YRI population **(Table 3.6).** The LWK and the YRI are the two population groups which are often used as the reference population for the Southern African population due to a paucity of allele frequency distribution for the local population groups. This is because these three populations are thought to share recent common ancestry with via the expansion of Bantu-speaking groups within the last 5000 years (May et al. 2013).

Several alleles in transporter genes were significantly different in frequency among the three populations, while none of the alleles in genes of the thymidine synthesis / metabolism pathway were significantly different in frequency among the three populations. This shows that the YRI and LWK are not always good proxies for genetic data in SA populations

Table 3.6 The comparison of minor allele frequencies of the SA population with MAF of other African populations. Significant p values (p<0.05) shown in red bold.

Gene	SNP	Position[b]	Minor allele frequencies[c]			p-value
			SA	LWK	YRI	
d4T transporter genes						
ABCC5	rs3749442	183942797	0.208	0.237	0.194	0.640
ABCG2	rs2725252	88140758	0.193	0.222	0.333	**0.002**
ABCG2	rs2622604	88157772	0.132	0.126	0.083	0.217
ABCG2	rs12505410	88109689	0.122	0.126	0.06	**0.039**
ABCG2	rs3114018	88143429	0.419	0.333	0.31	**0.035**
SLC28A3	rs4877847	84331502	0.470	0.475	0.463	0.9698
SLC28A3	rs7853758	84286011	0.5	0.303	0.366	**0.0001**
SLC28A1	rs2242046	84935498	0.025	0.061	0.000	**0.0008**
SLC28A1	rs2290272	84904200	0.242	0.172	0.171	0.059
SLC28A1	rs8187758	84905644	0.155	0.131	0.157	0.700
dNTP pool regulator genes						
MTHFR	rs1801131	11794419	0.144	0.187	0.120	0.159
MTHFR	rs1801133	11796321	0.070	0.071	0.106	0.284
RRM2	rs7574663	10119200	0.084	0.146	0.134	0.081
DHFR	rs1650723	80626211	0.062	0.071	0.046	0.565
RRM2B	rs16918482	102206460	0.064	0.076	0.065	0.889
RRM1	rs1465952	4093309	0.477	0.389	0.449	0.170
RRM1	rs11030918	4094257	0.20	0.167	0.181	0.632
RRM1	rs12806698	4094744	0.021	0.015	0.019	0.914
RRM1	rs1042927	4138699	0.181	0.263	0.190	0.079

Table 3.6 continued. The comparison of minor allele frequencies of the SA population with MAF of other African populations. Significant p values (p<0.05) shown in red bold.

Gene	SNP	Position[b]	Minor allele frequencies[c]			p-value
			SA	LWK	YRI	
TK2	rs11859474	66520637	0.263	0.293	0.278	0.518
TK2	rs2288399	66531589	0.152	0.081	0.106	0.058
SAMHD1	rs8124728	36892303	0.169	0.152	0.148	0.808
SAMHD1	rs1291142	36896959	0.231	0.237	0.231	0.984
SAMHD1	rs1891643	36941608	0.187	0.207	0.167	0.572
SLC19A1	rs1051266	45537880	0.311	0.308	0.329	0.892

[a] Chromosome numbers for SNPs.

[b] Position of the single nucleotide polymorphism on the respective chromosome.

[c] Minor allele frequencies for the LWK and YRI population from a 1000 genomes phase 3 database published on Ensembl.

(Red) p<0.05 significantly different from each other.

3.4.3 LD and Haplotype analyses in the SA cohort.

Linkage disequilibrium (LD) was analysed for seven genes with more than one SNP genotyped (*ABCG2*, *SLC28A1*, *SLC28A3*, *MTHFR*, *RRM1*, *TK2* and *SAMHD1*). SNPs in five of these genes showed strong linkage between the following polymorphisms:
RRM1: strong LD between rs1465952, rs11030918 and rs12806698 (Fig 3.4);
TK2: strong LD between rs3743712 and rs11859474 (D'= 1.000) (Fig 3.5);
SAMHD1: complete linkage between all three SNPs rs8124728, rs1291142 and rs1891643 (D'= 1.000) (Fig 3.6); *SLC28A1*: strong LD between s2290272 and rs8187758 (D'= 0.961) (Fig 3.7).

ABCG2: strong LD between rs12505410 and rs2725252 (D'= 1.000) (Fig 3.8).
The following SNPs from *MTHFR* rs1801131 and rs1801133 (D'= 0.009) and *SLC28A3* rs4877847 and rs7853758 (D'= 0.253) were not in significant LD with each other.

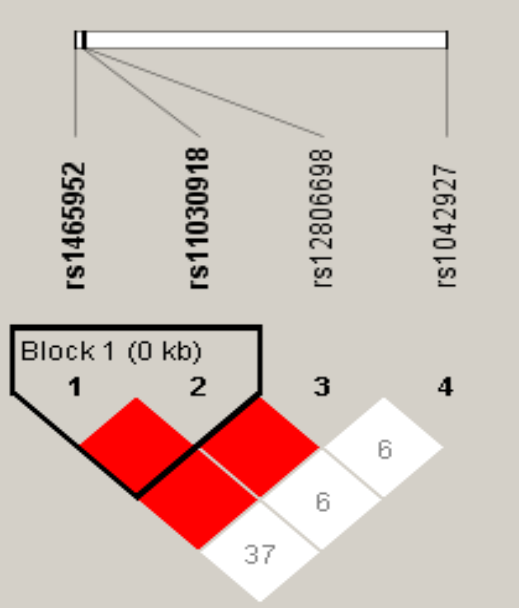

Figure 3.4 *RRM1* gene LD plot generated using Haploview. The red blocks on the LD plot depicted in this figure and those below indicate the strength of the LD, with the brighter the square the higher the LD and fainter red the lower the LD.

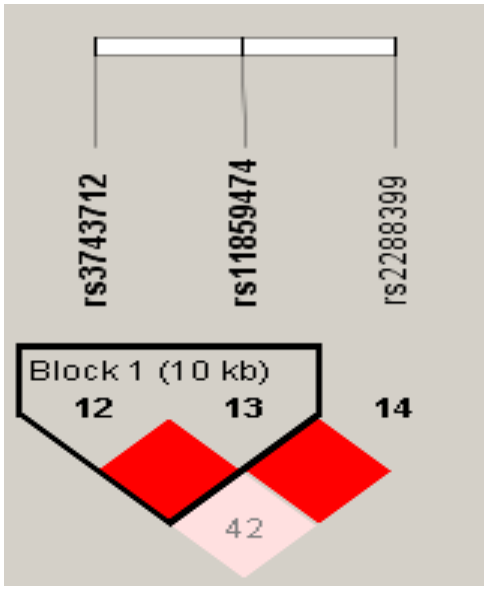

Figure 3.5 *TK2* gene LD plot generated using Haploview.

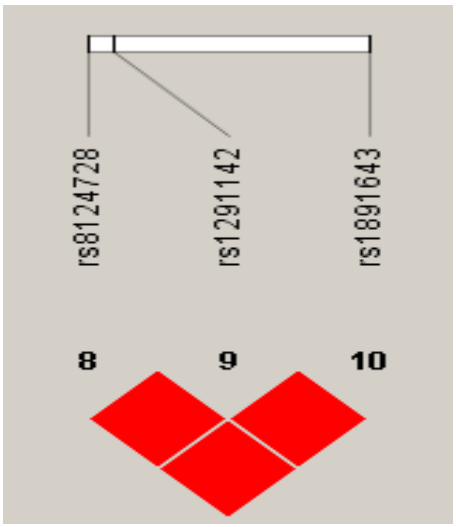

Figure 3.6 *SAMHD1* gene LD plot generated using Haploview.

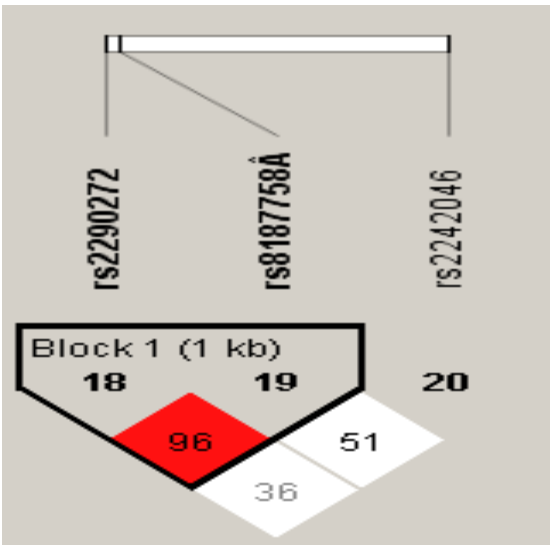

Figure 3.7 *SLC28A1* gene LD plot generated using Haploview.

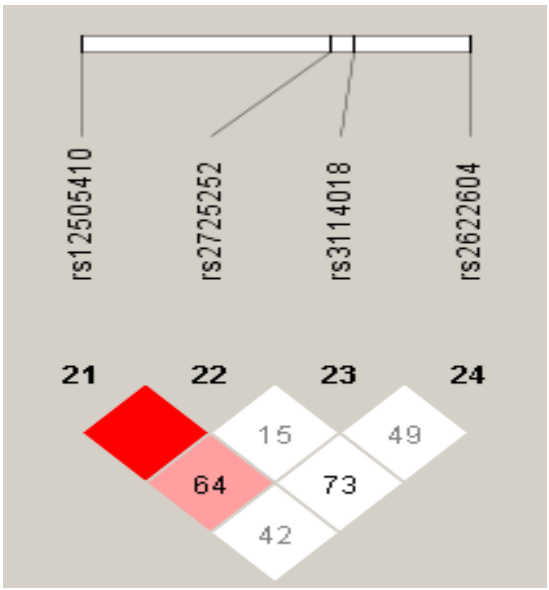

Figure 3.8 *ABCG2* gene LD plot generated using Haploview.

The frequencies of two and three-SNP haplotypes in the seven genes were calculated using the sliding window specification approach in PLINK. There were 42 haplotype combinations created from 7 genes (*ABCG2, SAMHD1, SLC28A1, SLC28A3, MTHFR, TK2* and *RRM1*), with 34 haplotypes derived from three-SNP combinations whereas 8 haplotypes were derived from two-SNP combinations **(Table 3.7).**

Table 3.7 Haplotype frequencies in the SA cohort.

Gene	Haplotype	Frequency	loci in the haplotype
MTHFR	GA	0.011	rs1801131|rs1801133
	TA	0.059	rs1801131|rs1801133
	GG	0.133	rs1801131|rs1801133
	TG	0.797	rs1801131|rs1801133
ABCG2	TAC	0.101	rs12505410|rs2725252|rs3114018
	GCC	0.099	rs12505410|rs2725252|rs3114018
	TCC	0.219	rs12505410|rs2725252|rs3114018
	TAA	0.092	rs12505410|rs2725252|rs3114018
	GCA	0.024	rs12505410|rs2725252|rs3114018
	TCA	0.465	rs12505410|rs2725252|rs3114018
	CCT	0.022	rs2725252|rs3114018|rs2622604
	CAT	0.104	rs2725252|rs3114018|rs2622604
	ACC	0.095	rs2725252|rs3114018|rs2622604
	CCC	0.297	rs2725252|rs3114018|rs2622604
	AAC	0.094	rs2725252|rs3114018|rs2622604
	CAC	0.383	rs2725252|rs3114018|rs2622604
SLC28A3	AC	0.174	rs7853758|rs4877847
	GC	0.295	rs7853758|rs4877847
	AA	0.323	rs7853758|rs4877847
	GA	0.208	rs7853758|rs4877847
RRM1	ACA	0.021	rs1465952|rs11030918|rs12806698
	ACC	0.18	rs1465952|rs11030918|rs12806698
	ATC	0.277	rs1465952|rs11030918|rs12806698
	GTC	0.522	rs1465952|rs11030918|rs12806698
	CCC	0.041	rs11030918|rs12806698|rs1042927
	TCC	0.135	rs11030918|rs12806698|rs1042927
	CAA	0.016	rs11030918|rs12806698|rs1042927
	CCA	0.139	rs11030918|rs12806698|rs1042927
	TCA	0.664	rs11030918|rs12806698|rs1042927
SLC28A1	GCA	0.012	rs2290272|rs8187758Â |rs2242046
	AAG	0.149	rs2290272|rs8187758Â |rs2242046
	ACG	0.079	rs2290272|rs8187758Â |rs2242046
	GCG	0.743	rs2290272|rs8187758Â |rs2242046

Table 3.7 continued. Haplotype frequencies in the SA cohort.

Gene	Haplotype	Frequency	loci in the haplotype
TK2	GGG	0.11	rs3743712|rs11859474|rs2288399
	AGG	0.042	rs3743712|rs11859474|rs2288399
	GAA	0.263	rs3743712|rs11859474|rs2288399
	GGA	0.017	rs3743712|rs11859474|rs2288399
	AGA	0.567	rs3743712|rs11859474|rs2288399
SAMHD1	GGG	0.187	rs8124728|rs1291142|rs1891643
	GGA	0.045	rs8124728|rs1291142|rs1891643
	AAA	0.169	rs8124728|rs1291142|rs1891643
	GAA	0.598	rs8124728|rs1291142|rs1891643

3.4.4 SNP Genotype frequencies in SA population

The determination of the genotypes for all the 26 SNPs in 243 DNA samples in this study allowed for the subsequent determination of the genotype frequencies for the Southern African population. The Chi-square test of observed versus expected genotype frequencies was conducted and all the genotypes were found to be in Hardy-Weinberg equilibrium **(Appendix B).**

3.5 Associations between SNPs and sensory neuropathy.

A total of 26 SNPs, including ten SNPs from four drug transporter genes and 16 SNPs from eight dNTP pool regulator genes were used to analyse association with sensory neuropathy (SN) using univariate and multivariate analysis methods.

3.5.1 Univariate SNP association with SN

The single SNP univariate analyses for association with SN was done using PLINK using the allelic and three genotype models (genotypic, dominant and recessive).

Only one SNP of the *SLC28A1* transporter gene (**709C>A; rs8187758**) was found to be significantly associated with sensory neuropathy. This SNP was significantly associated with sensory neuropathy in the genotypic model test PEMP1<0.05 at p=**0.026**, but not significant for the PEMP2<0.05 correction for multiple testing. No association was found for all the other models tested (allelic, dominant and recessive) **(Table 3.8, highlighted**.)

Table 3.8 Univariate analyses of association of alleles and genotypes with sensory neuropathy. Significant p-values (p<0.05) highlighted in red.

Univariate analyses for association of alleles and genotypes with sensory neuropathy.									
Gene	SNP	Allelic		Genotypic		Dominant		Recessive	
		P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}
Drug transporters genes									
ABCC5	rs3749442	0.255	0.998	0.351	1.000	0.425	1.000	0.194	0.955
ABCG2	rs12505410	0.159	0.976	0.327	1.000	0.136	0.978	0.876	1.000
	rs2725252	0.908	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	rs3114018	0.284	1.000	0.559	1.000	0.398	1.000	0.314	0.999
	rs2622604	0.535	1.000	0.867	1.000	0.507	1.000	0.791	1.000
SLC28A3	rs7853758	0.782	1.000	0.913	1.000	0.655	1.000	1.000	1.000
	rs4877847	0.748	1.000	0.428	1.000	0.350	1.000	0.641	1.000
SLC28A1	rs2290272	0.174	0.991	0.154	0.981	0.082	0.855	0.810	1.000
	rs8187758	0.226	0.998	**0.026**	0.458	0.061	0.731	0.193	0.978
	rs2242046	0.235	0.997	0.236	0.998	0.238	0.997	1.000	1.000

(Red) p<0.05 Significant association with sensory neuropathy.

Table 3.8 continued. Univariate analyses of association of alleles and genotypes with sensory neuropathy.

dNTP pool regulators genes									
Gene	SNP	Allelic		Genotypic		Dominant		Recessive	
		P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}
MTHFR	rs1801131	0.239	0.997	0.376	1.000	0.383	1.000	0.329	0.999
	rs1801133	1.000	1.000	0.917	1.000	1.000	1.000	1.000	1.000
RRM2	rs7574663	0.629	1.000	0.861	1.000	0.604	1.000	1.000	1.000
DHFR	rs1650723	0.847	1.000	0.846	1.000	0.846	1.000	1.000	1.000
RRM2B	rs16918482	0.722	1.000	0.919	1.000	0.843	1.000	0.78	1.000
RRM1	rs1465952	0.930	1.000	0.864	1.000	0.950	1.000	0.716	1.000
	rs11030918	0.420	1.000	0.415	1.000	0.688	1.000	0.141	0.941
	rs12806698	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	rs1042927	0.701	1.000	0.829	1.000	0.685	1.000	1.000	1.000
TK2	rs3743712	0.527	1.000	0.493	1.000	0.315	1.000	1.000	1.000
	rs11859474	0.920	1.000	0.424	1.000	0.518	1.000	0.484	1.000
	rs2288399	0.905	1.000	0.774	1.000	1.000	1.000	0.521	1.000
SAMHD1	rs8124728	0.720	1.000	**0.055**	0.746	0.263	0.999	0.108	0.863
	rs1291142	0.332	1.000	0.246	1.000	0.170	0.987	0.684	1.000
	rs1891643	1.000	1.000	0.372	1.000	0.685	1.000	0.301	0.997
SLC19A1	rs1051266	0.186	0.991	0.328	1.000	0.157	0.972	0.815	1.000

3.5.2 Multivariate analysis of SNP association with SN.

Next multivariate analysis using logistic regression was performed taking into account two demographic covariates (age and height) simultaneously. All SNPs were analysed during multivariate analysis, regardless of significance status during univariate testing. The analysis deployed four tests which were the allelic test and three genotype tests (dominant, recessive and genotypic/co-dominant).

Alleles and or genotypes of four SNPs from the d4T metabolism / dNTP regulation pathways were associated with SN in the multivariate analysis after correcting for age and height, but before correction for multiple testing **(Table 3.9)** as follows:

***MTHFR* rs1801131:** this missense variant was associated with SN after correcting for age and height, in genotypic and recessive models, with point-wise estimates (PEMP1), p=**0.001** and p=**0.001** respectively. The corrected (PEMP2) values were not significant.

***RRM2B* rs16918482 :** this 3' UTR variant was found to be associated with SN after correcting for age and height, in genotypic and recessive models, with point-wise estimates (PEMP1), p=**0.001** and p=**0.00001** respectively. The correction values (PEMP2) were not significant for all the tests.

***SAMHD1* rs8124728:** this 3' UTR variant was found to be associated with SN after correcting for age and height, in genotypic and recessive model with the point-wise estimates (PEMP1), p=**0.017** and p=**0.015** respectively. The correction values (PEMP2) were not significant for all the tests.

***SAMHD1* rs1891643:** this intron variant was found to be associated with SN after correcting for age and height, in the recessive test for the point-wise estimate (PEMP1), p= **0.027** with correction (PEMP2) not significant for all the tests.

None of these SNPs or genotypes were significantly associated with SN in a multivariate model after correction for multiple testing (all PEMP2 values >0.05, Table 3.9).

Table 3.9 Multivariate analysis of SNP association with d4T-induced sensory neuropathy; age and height were used as covariates. Significant p-values (p<0.05) highlighted in red.

Multivariate logistic regression analysis for association with d4T-induced sensory neuropathy.									
Gene	SNP	Allelic		Genotypic		Dominant		Recessive	
		P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}
Drug transporter genes									
ABCC5	rs3749442	0.370	1.000	0.262	0.986	0.551	1.000	0.3000	0.989
ABCG2	rs12505410	0.159	0.984	0.698	1.000	0.149	0.974	0.746	1.000
	rs2725252	0.885	1.000	0.829	1.000	0.957	1.000	0.866	1.000
	rs3114018	0.272	0.999	0.317	1.000	0.284	1.000	0.508	1.000
	rs2622604	0.392	1.000	0.319	1.000	0.492	1.000	0.330	1.000
SLC28A3	rs7853758	0.606	1.000	0.609	1.000	0.544	1.000	0.822	1.000
	rs4877847	0.468	1.000	0.521	1.000	0.213	0.997	0.913	1.000
SLC28A1	rs2290272	0.28	1.000	0.892	1.000	0.126	0.961	0.558	1.000
	rs8187758	0.410	1.000	0.171	0.975	0.144	1.000	0.073	0.923
	rs2242046	0.240	1.000	1.000	1.000	0.242	1.000	1.000	1.000

(Red Bold) p<0.05 Significant association with sensory neuropathy.

Table 3.9 continued. Multivariate analysis of SNP association with d4T-induced sensory neuropathy; age and height were used as covariates. Significant p-values (p<0.05) highlighted in red.

dNTP pool regulator genes									
Gene	SNP	Allelic		Genotypic		Dominant		Recessive	
		P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}
MTHFR	rs1801131	0.134	0.968	**0.001**	0.852	0.239	0.998	**0.002**	0.875
	rs1801133	0.996	1.000	0.438	1.000	0.853	1.000	0.443	1.000
RRM2	rs7574663	0.434	1.000	0.477	1.000	0.389	1.000	0.40	1.000
DHFR	rs1650723	0.801	1.000	1.000	1.000	0.801	1.000	1.000	1.000
RRM2B	rs16918482	0.881	1.000	**0.001**	1.000	0.940	1.000	**0.00001**	1.000
RRM1	rs1465952	0.936	1.000	0.973	1.000	0.552	1.000	0.619	1.000
	rs11030918	0.543	1.000	0.294	0.996	0.785	1.000	0.301	0.996
	rs12806698	0.517	1.000	1.000	1.000	0.517	1.000	1.000	1.000
	rs1042927	0.800	1.000	0.706	1.000	0.723	1.000	0.717	1.000
TK2	rs3743712	0.970	1.000	0.764	1.000	0.631	1.000	0.493	1.000
	rs11859474	0.677	1.000	0.302	0.997	0.905	1.000	0.241	0.985
	rs2288399	0.911	1.000	0.734	1.000	0.984	1.000	0.730	1.000
SAMHD1	rs8124728	0.739	1.000	**0.017**	0.495	0.661	1.000	**0.015**	0.350
	rs1291142	0.455	1.000	0.718	1.000	0.237	0.999	0.562	1.000
	rs1891643	0.672	1.000	0.619	0.698	0.862	1.000	**0.027**	0.619
SLC19A1	rs1051266	0.079	0.858	1.000	0.984	0.067	0.806	0.482	1.000

(Red Bold) p<0.05 Significant association with sensory neuropathy.

3.5.3 Analysis of associations between haplotypes and SN.

Haplotype analysis was performed for all the genes with more than one SNP, using the sliding window approach in PLINK. Two and three-SNP haplotype combinations were analysed. The haplotype analysis for multivariate was performed correcting for age and height.

In univariate analysis, no haplotype was associated with SN **(Table 3.10).** One haplotype was associated with SN in the multivariate analysis **(Table 3.10).** This was the three-SNP **(rs8124728|rs1291142|rs1891643)** haplotype combination "GGA" of the *SAMHD1* gene, which was found to be significant in the multivariate analysis, (PEMP1, p=**0.021**). However, this haplotype was not significantly associated with SN after correction for multiple testing (PEMP1>0.05; **Table 3.10**).

Table 3.10 The univariate and multivariate analysis of associations between haplotypes and sensory neuropathy. Significant p values<0.05 shown in red.

Gene	Haplo-type	SNPs in the haplotype	Frequency SN+ Cases N=143	Frequency SN- Controls N=120	Univariate analysis		Multivariate analysis	
					PEMP1	PEMP2	PEMP1	PEMP2
MTHFR	GA	rs1801131|rs1801133	0.001	0.012	0.565	1.000	0.575	1.000
	TA	rs1801131|rs1801133	0.061	0.057	0.802	1.000	0.919	1.000
	GG	rs1801131|rs1801133	0.117	0.153	0.227	1.000	0.141	1.000
	TG	rs1801131|rs1801133	0.812	0.778	0.342	1.000	0.227	1.000
ABCG2	TAC	rs12505410|rs2725252|rs3114018	0.105	0.097	0.759	1.000	0.517	1.000
	GCC	rs12505410|rs2725252|rs3114018	0.110	0.084	0.295	1.000	0.273	1.000
	TCC	rs12505410|rs2725252|rs3114018	0.223	0.215	0.824	1.000	0.993	1.000
	TAA	rs12505410|rs2725252|rs3114018	0.085	0.100	0.512	1.000	0.617	1.000
	GCA	rs12505410|rs2725252|rs3114018	0.031	0.016	0.231	1.000	0.327	1.000
	TCA	rs12505410|rs2725252|rs3114018	0.445	0.488	0.284	1.000	0.248	1.000
	CCT	rs2725252|rs3114018|rs2622604	0.024	0.020	0.697	1.000	0.592	1.000
	CAT	rs2725252|rs3114018|rs2622604	0.113	0.095	0.466	1.000	0.412	1.000
	ACC	rs2725252|rs3114018|rs2622604	0.099	0.091	0.746	1.000	0.487	1.000
	CCC	rs2725252|rs3114018|rs2622604	0.314	0.280	0.362	1.000	0.521	1.000
	AAC	rs2725252|rs3114018|rs2622604	0.088	0.101	0.579	1.000	0.611	1.000
	CAC	rs2725252|rs3114018|rs2622604	0.362	0.414	0.196	1.000	0.146	1.000

Table 3.10 continued. The univariate and multivariate analysis of associations between haplotypes and sensory neuropathy. Significant p values<0.05 shown in red.

Gene	Haplo-type	SNPs in the haplotype	Frequency SN+ Cases N=143	Frequency SN- Controls N=120	Univariate analysis		Multivariate analysis	
					PEMP1	PEMP2	PEMP1	PEMP2
SLC28A3	GC	rs7853758|rs4877847	0.295	0.294	0.974	1.000	0.768	1.000
	AA	rs7853758|rs4877847	0.340	0.303	0.377	1.000	0.192	0.999
	GA	rs7853758|rs4877847	0.198	0.220	0.501	1.000	0.532	1.000
RRM1	ACA	rs1465952|rs11030918|rs12806698	0.022	0.019	0.995	1.000	0.516	1.000
	ACC	rs1465952|rs11030918|rs12806698	0.193	0.165	0.430	1.000	0.653	1.000
	ATC	rs1465952|rs11030918|rs12806698	0.260	0.298	0.354	1.000	0.630	1.000
	GTC	rs1465952|rs11030918|rs12806698	0.525	0.518	0.920	1.000	0.929	1.000
	CCC	rs11030918|rs12806698|rs1042927	0.041	0.041	0.974	1.000	0.754	1.000
	TCC	rs11030918|rs12806698|rs1042927	0.131	0.143	0.682	1.000	0.867	1.000
	CAA	rs11030918|rs12806698|rs1042927	0.018	0.013	0.607	1.000	0.524	1.000
	CCA	rs11030918|rs12806698|rs1042927	0.151	0.125	0.366	1.000	0.536	1.000
	TCA	rs11030918|rs12806698|rs1042927	0.658	0.679	0.627	1.000	0.665	1.000
SLC28A1	GCA	rs2290272|rs8187758Â |rs2242046	0.021	0.002	0.090	0.996	0.1	0.997
	AAG	rs2290272|rs8187758Â |rs2242046	0.126	0.183	0.082	0.943	0.204	1.000
	ACG	rs2290272|rs8187758Â |rs2242046	0.080	0.079	0.964	1.000	0.950	1.000
	GCG	rs2290272|rs8187758Â |rs2242046	0.773	0.737	0.455	1.000	0.640	1.000
TK2	GGG	rs3743712|rs11859474|rs2288399	0.104	0.119	0.603	1.000	0.808	1.000
	AGG	rs3743712|rs11859474|rs2288399	0.046	0.037	0.752	1.000	0.867	1.000
	GAA	rs3743712|rs11859474|rs2288399	0.261	0.266	0.882	1.000	0.680	1.000
	GGA	rs3743712|rs11859474|rs2288399	0.012	0.024	0.348	1.000	0.416	1.000
	AGA	rs3743712|rs11859474|rs2288399	0.578	0.554	0.607	1.000	0.959	1.000
SAMHD1	GGG	rs8124728|rs1291142|rs1891643	0.187	0.188	0.978	1.000	0.670	1.000
	GGA	rs8124728|rs1291142|rs1891643	0.030	0.064	0.066	0.907	**0.021**	0.566
	AAA	rs8124728|rs1291142|rs1891643	0.175	0.163	0.684	1.000	0.755	1.000
	GAA	rs8124728|rs1291142|rs1891643	0.608	0.586	0.592	1.000	0.394	1.000

3.6 Association between mtDNA copy number and sensory neuropathy

3.6.1 Primer optimization using regular PCR

Two sets of PCR primers for the *GAPDH* gene and the two sets of PCR primers for the *mt-ND1* gene were optimized initially using the gradient PCR to determine the optimal annealing temperature before using them in the qPCR. Both sets of *GAPDH* and *mt-ND1* primer pairs worked at all temperatures included in the gradient (see **Figure 3.9** for *GAPDH* primer pairs and **Figure 3.10** for *mt-ND1* primer pairs) and therefore were tested in the qPCR assay.

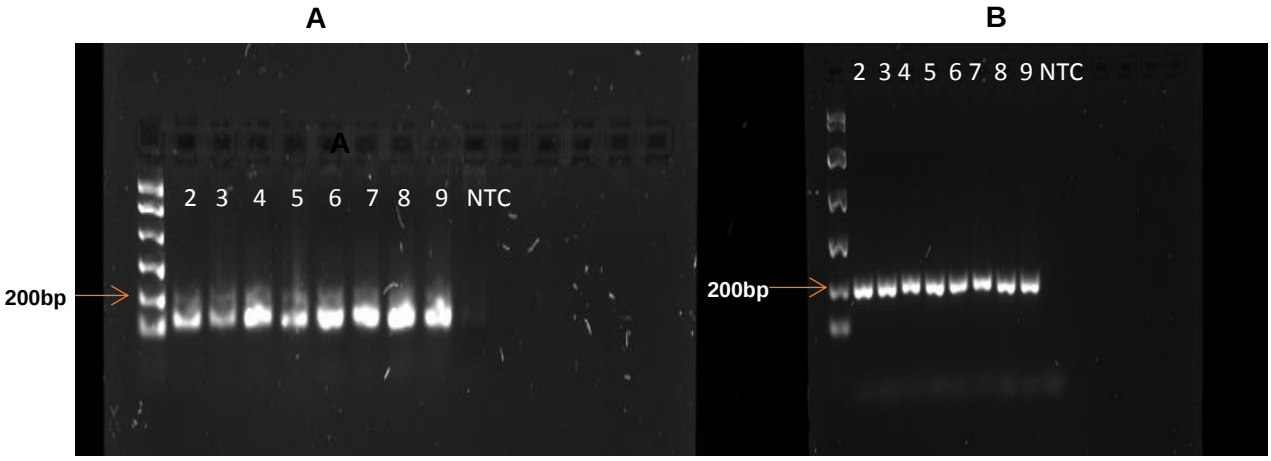

Figure 3.9: Agarose gel 1% (w/v) electrophoresis images of the PCR products for *GAPDH* primer pairs.

a) *GAPDH* primer pair 2 at temperature gradient between 45°C and 50°C (lanes 2-9). Clear NTC lane indicated no contamination. Expected product size was 197 bp.

b) *GAPDH* primer pair 1 at temperature gradient between 48°C and 53°C (lanes 2-9). Clear NTC lane indicated no contamination. Expected product size was 208 bp. Kapa DNA Ladder with the following sizes 100+200+400+800+1600+4000+8000 bp was used and the last lane was a no template control (NTC).

The *mt-ND1* primer pairs were also demonstrated to be specific with all the selected PCR gradient temperatures.

Both the primer pairs were tested in the qPCR assay.

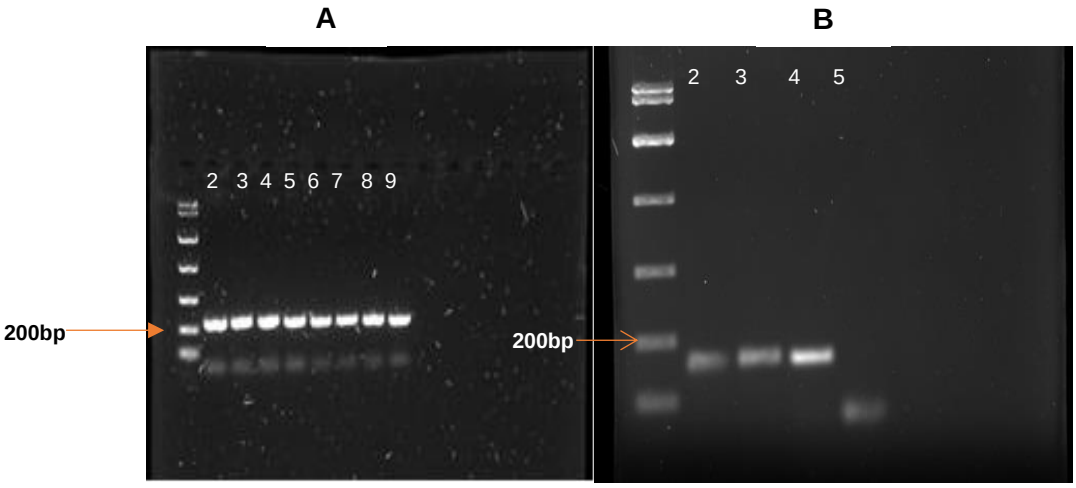

Figure 3.10: Agarose gel 1% (w/v) electrophoresis images of the PCR products for *mt-ND1* primer pairs.

a) *mt-ND1* primer pair 1 at temperature gradient between 47°C and 54°C from (lanes 2-9). Expected product size was 243 bp.

b) *mt-ND1* primer pair 2 at gradient temperature used was between 58°C and 62 °C from (lanes 2-4). Expected product size was 154 bp.

Kapa DNA Ladder with the following sizes 100+200+400+800+1600+4000+8000 bp was used and the last lane was a no template control (NTC).

3.6.2 qPCR optimization

3.6.2.1 Standard curves

In order to calculate the PCR efficiency of each primer pair, a four point series of tenfold dilutions of the HEK293 DNA was used.

The four DNA dilutions were used to amplify the target gene *mt-ND1* and the reference gene *GAPDH* using duplicate reactions. Post the real-time PCR the CT values were used to construct a standard curve using the second derivative analysis method on the Roche Light cycler 480 software.

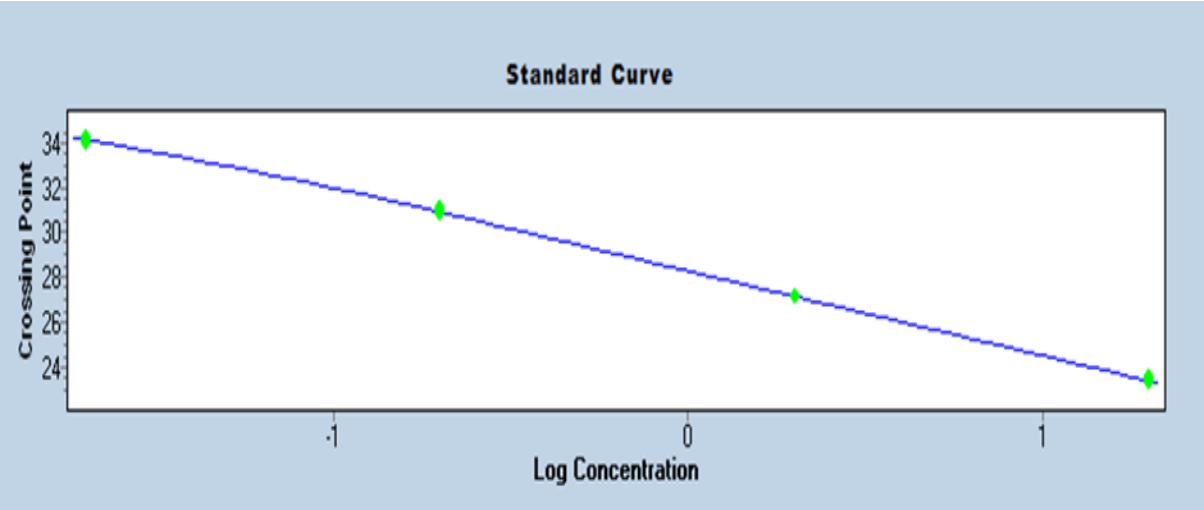

Figure 3.11 Standard curve graph used to determine the PCR efficiencies of *GAPDH* primers.

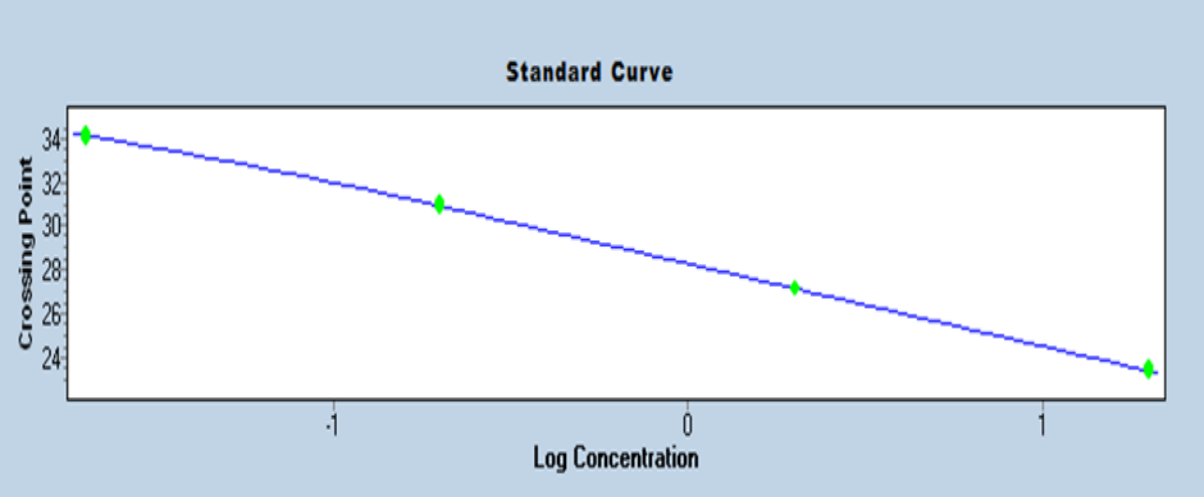

Figure 3.12 Standard curve graph used to determine the PCR efficiencies of *mt-DN1* primers.

The DNA amount, primer concentrations and annealing temperatures were altered in order to attain similar amplification efficiencies for the target and the reference gene. The PCR efficiencies were calculated from standard curve slopes using the Life Cycler software. The determination of the corresponding real-time PCR efficiency (*E*) of one cycle in the exponential phase was calculated using the equation: $E=10^{[-1/slope]}$ **(**and the PCR efficiencies for all the four primer pairs used were within the range between 90 % and 105 % **(Table 3.11.).**

Table 3.11 qPCR PCR efficiencies for the primers selected for optimization.

	REFERENCE GENE		**TARGET GENE**	
Primer name	*GAPDH1*	*GAPDH2*	*mt-ND1A*	*mt-ND1B*
Slope	-3.477	-3.229	-3.550	-3.230
Efficiency (%)	**93.9**	**104**	**91.3**	**104**
Y-Intercept	25.79	26.04	24.21	20.68
Ta (°C)	50 °C	50 °C	50 °C	58 °C
Error	0.026	0.002	0.056	0.131

3.6.2.2 Melting curves

The post-amplification melting analysis was performed in order to check the specificity of the primers. Following the post amplification melting curve analysis for the *GAPDH* primer pair 1 **(Figure 3.13)** one peak was observed showing that it was specific to the region of interest, whereas primer pair 2 **(Figure 3.14)** showed two peaks on its melting curve for the gradient temperature range tested indicating non-specificity. Hence for the reference gene *GAPDH* primer pair 1 was chosen as the primer pair to be used in the mtDNA copy number quantification assay. For the target gene both the primer pairs for the *mt-ND1* gene have shown to be specific with no secondary peaks observed on the melting curve, both the primer pairs for the *mt-ND1* target gene were both eligible for selection, as a result primer pair 2 **(Figure 3.15)** was chosen to be used in the assay for mtDNA copy number quantification assay.

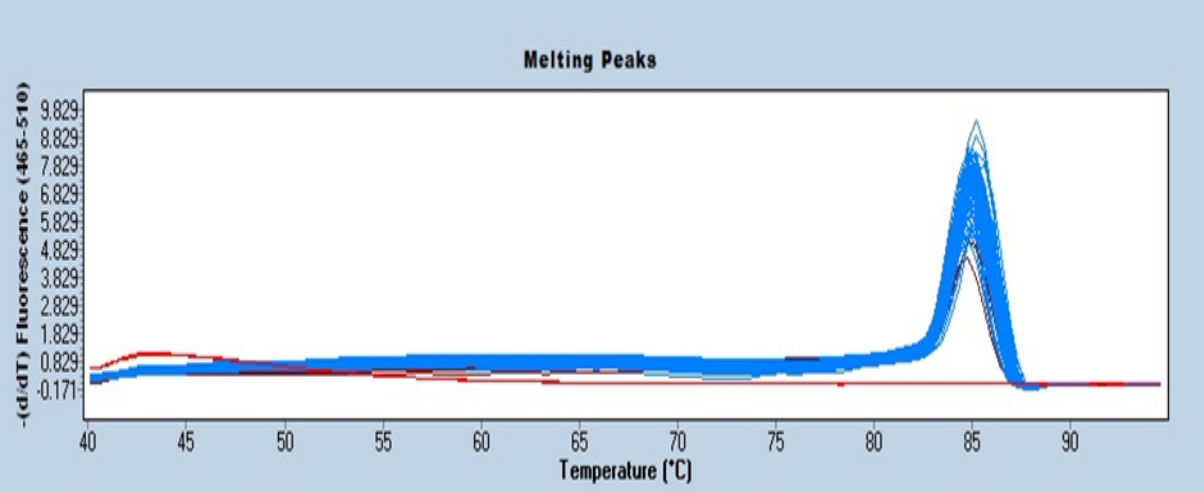

Figure 3.13 A melting curve analysis test for *GAPDH* primer Pair 1.

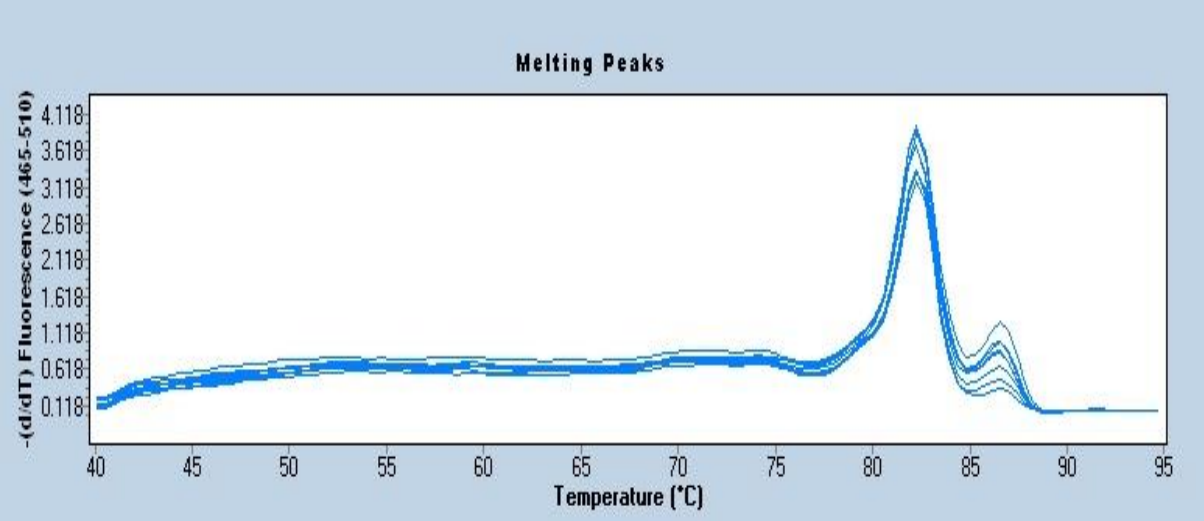

Figure 3.14 A melting curve analysis test for *GAPDH* primer pair 2.

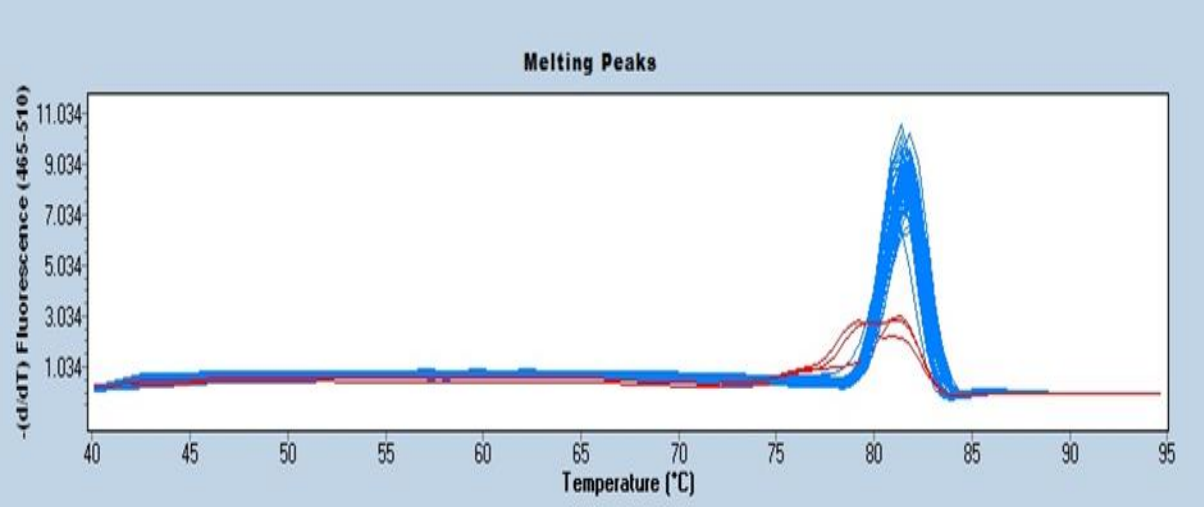

Figure 3.15 A melting curve analysis test for *mt-ND1* primer pair 2.

3.6.2.3. Validation of qPCR assay for mtDNA copy number calculation

A validation procedure for the mtDNA copy number quantification assay was used to confirm the accuracy and reproducibility of the qPCR assay before the cohort samples were tested. HEK293 cells were treated with EtBr for 6 days and DNA was extracted from cells preserved at the day 0 (pre-treatment), days 3 and 6 time points. DNA concentrations were 163.3 ng/µl and for the following days, day 3 and 6 post-treatment with EB were 236.9 ng/µl and 94.2 ng/µl respectively. The DNA was diluted to a working concentration of 20 ng/µl.

The DNA from day 0(untreated) , day 3 (treated with EtBr) and day 6 (treated with EtBr) were amplified using the *GAPDH* primer pair 1 and *mt-ND1* primer pair 2; this was carried out in triplicates and the Cts were averaged and compared for statistical differences between the days. To calculate the assay precision the Pfaffl equation for fold change calculations was used to measure the inter-treatment change of the Cts between the control group (pre-treatment) and the treated groups (day 3 and 6).
Averages of Ct values for replicate samples were calculated as shown in **Table 3.12**. The Cts values of the target gene increased from day 0 to day 3, and from day 3 to day 6, indicating decreased mtDNA, and the Ct values of the reference gene remained fairly constant for all time points.

Table 3.12: **qPCR threshold cycle (Ct) values obtained using the DNA of HEK293 cells at three time points.**

Days	Target (*mt-ND1*)			Reference (*GAPDH*)		
	replicate Ct Values		Average Ct	replicate Ct Values		Average Ct
0	14.28	16.00	**15.14**	23.08	23.51	**23.3**
3	16.82	17.52	**17.17**	22.89	22.58	**22.7**
6	19.17	19.87	**19.52**	22.91	22.87	**22.9**

Table 3.13 shows the fold changes from day 0 to day 6. The Livak equation ($2^{-\Delta\Delta CT}$) was used for calculation of fold change of nDNA: mtDNA ratio between EtBr treated and untreated HEK293 cells.

The mtDNA to nDNA ratio varied across the respective days. The EtBr was effective in inducing mtDNA copy number depletion in HEK293 cells. Day 3 showed a 6-fold decrease in *mt-ND1* gene compared to day 0, while on day 6 the fold change increased to 27.67-fold compared to day 0.

These results supported the expected decrease over time in mtDNA in cells treated with EtBr and proved that the qPCR assay was working correctly to detect mtDNA change.

Table 3.13: The fold change ratios of *mt-ND1* relative to *GAPDH* in HEK293 cells.

Days	ΔΔCt	$2^{-\Delta\Delta Ct}$ (fold change) **(Treated compared to untreated)**
(0-3)	2.59	6.02
(3-6)	2.195	4.57
(0-6)	4.79	27.67

3.6.3 Associations between mtDNA copy number and sensory neuropathy.

We performed the mtDNA: nDNA qPCR assay in 263 samples in duplicate wells. After QC, 27 samples were excluded and 236 samples remained for further analysis leaving 130 cases and 106 controls. mtDNA copy number was compared between HIV patients with SN and those without SN using the ΔΔCt method **(Table 3.14).**

Table 3.14 Fold change calculations for the cases and controls.

Subjects	Mean Cts *GAPDH*	Mean Cts *mt-ND1*	ΔCt	ΔΔCt	$2^{-\Delta\Delta Ct}$ (fold change)
Cases (with SN)	21.022	13.814	7.208	**-0.184**	**1.136**
Controls (without SN)	21.132	13.740	7.392		

The Man-Whitney U-test was used to compare differences in the mean mtDNA ratios of the cases and controls peripheral blood mononuclear cells (PBMCs) (1.445 vs. 1.268) **(p=0.159; Figure 3.16)**.

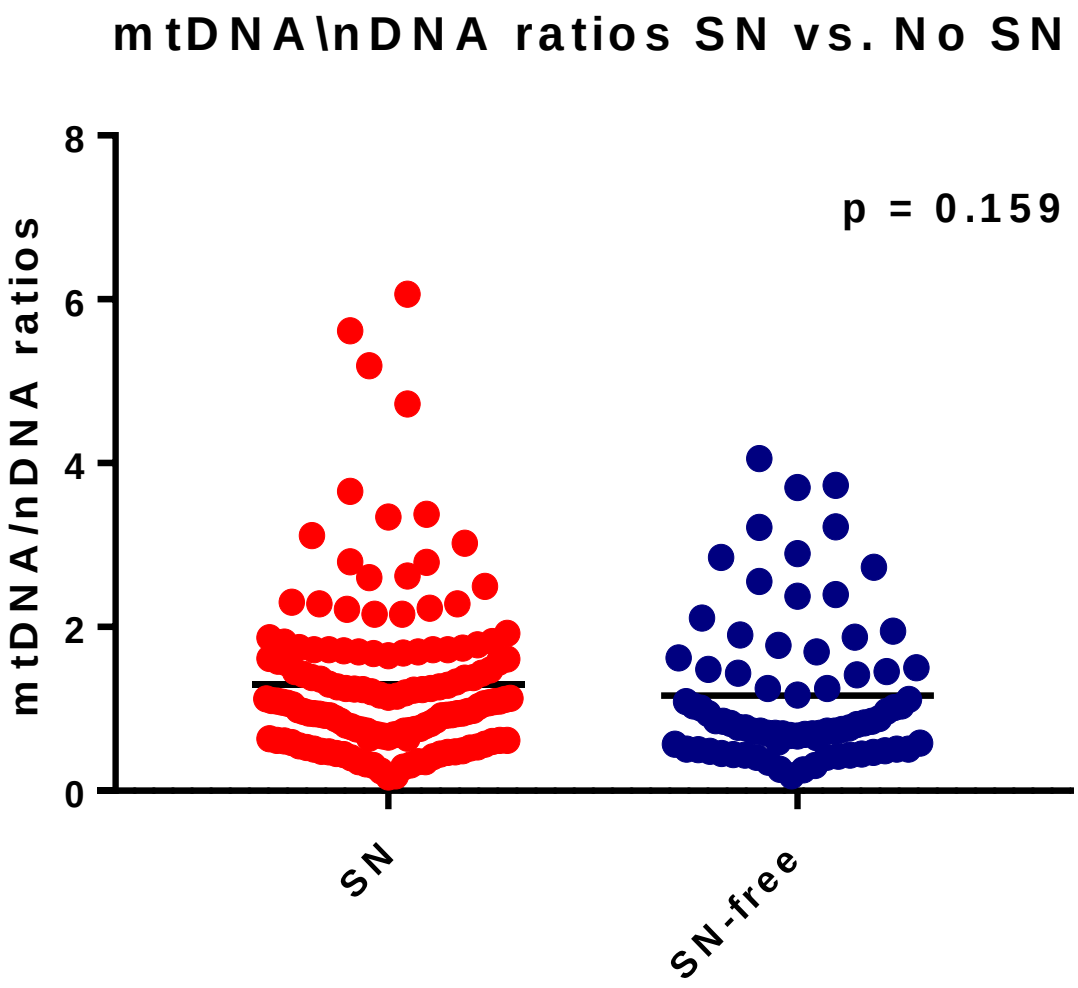

Figure 3.16 the comparison of mtDNA\nDNA Ct ratios between SN and without SN samples, using the non-parametric Man-Whitney test.

3.7 Associations between SNPs and mtDNA copy number

A total of 26 SNPs were examined for association with mtDNA copy number both in a univariate and multivariate analysis method. The mtDNA copy number phenotype was

expressed as Ct ratios between *mt-ND1* (mtDNA) and *GAPDH* (nDNA) (mtDNA/nDNA) and as a result the phenotype was analysed for association with the SNPs as a continuous or quantitative trait using PLINK.

3.7.1 Univariate analysis for association with mtDNA copy number

The single SNP univariate analyses for association with mtDNA copy number (mtDNA/nDNA ratio) was done using PLINK deploying four association tests, the allelic test and three genotype tests (dominant, recessive and co-dominant/genotypic). All 26 SNPs used in the univariate analysis of association with mtDNA copy number did not attain the threshold significance level for the point-wise estimates (PEMP1) and remained non-significant after correction for multiple testing (PEMP2) in all four models tested (allelic, genotypic, dominant and recessive) **(see Table 3.15).**

Table 3.15 Univariate analyses for association with mtDNA copy number.

Univariate analyses for association with mtDNA copy number.									
Gene	**SNP**	**Allelic**		**Genotypic**		**Dominant**		**Recessive**	
		P_{EMP1}	**P_{EMP2}**	**P_{EMP1}**	**P_{EMP2}**	**P_{EMP1}**	**P_{EMP2}**	**P_{EMP1}**	**P_{EMP2}**
Drug transporter genes									
ABCC5	rs3749442	0.445	1.000	0.448	1.000	0.447	1.000	0.446	1.000
ABCG2	rs12505410	0.277	1.000	0.276	1.000	0.279	1.000	0.280	1.000
	rs2725252	0.897	1.000	0.897	1.000	0.896	1.000	0.897	1.000
	rs3114018	0.323	1.000	0.322	1.000	0.322	1.000	0.320	1.000
	rs2622604	0.305	1.000	0.306	1.000	0.305	1.000	0.304	1.000
SLC28A3	rs7853758	0.281	0.999	0.277	0.999	0.281	0.999	0.277	0.999
	rs4877847	0.220	0.996	0.218	0.996	0.219	0.996	0.220	0.996
SLC28A1	rs2290272	0.163	0.982	0.161	0.982	0.160	0.982	0.161	0.981
	rs8187758	0.313	1.000	0.313	1.000	0.313	1.000	0.313	1.000
	rs2242046	0.813	1.000	0.811	1.000	0.810	1.000	0.809	1.000
dNTP pool regulator genes									
MTHFR	rs1801131	0.619	1.000	0.625	1.000	0.621	1.000	0.622	1.000
MTHFR	rs1801133	0.522	1.000	0.522	1.000	0.525	1.000	0.524	1.000
RRM2	rs7574663	0.989	1.000	0.988	1.000	0.988	1.000	0.988	1.000
DHFR	rs1650723	0.349	1.000	0.349	1.000	0.352	1.000	0.345	1.000

Table 3.15 continued. Univariate analyses for association with mtDNA copy number.

Gene	SNP	Allelic		Genotypic		Dominant		Recessive	
		P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}
dNTP pool regulator genes									
RRM2B	rs16918482	0.332	1.000	0.333	1.000	0.334	1.000	0.331	1.000
RRM1	rs1465952	0.583	1.000	0.58	1.000	0.58	1.000	0.580	1.000
	rs11030918	0.458	1.000	0.453	1.000	0.454	1.000	0.457	1.000
	rs12806698	0.964	1.000	0.963	1.000	0.964	1.000	0.963	1.000
	rs1042927	0.346	1.000	0.344	1.000	0.345	1.000	0.346	1.000
TK2	rs3743712	0.592	1.000	0.590	1.000	0.588	1.000	0.593	1.000
	rs11859474	0.673	1.000	0.671	1.000	0.671	1.000	0.673	1.000
	rs2288399	0.913	1.000	0.912	1.000	0.912	1.000	0.912	1.000
SAMHD1	rs8124728	0.629	1.000	0.632	1.000	0.632	1.000	0.629	1.000
	rs1291142	0.757	1.000	0.760	1.000	0.758	1.000	0.758	1.000
	rs1891643	0.659	1.000	0.661	1.000	0.662	1.000	0.659	1.000
SLC19A1	rs1051266	0.661	1.000	0.661	1.000	0.662	1.000	0.661	1.000

(Red) $p<0.05$ Significant association with mtDNA copy number.

3.8 SNP association with the CD4+ T cell count

A total of 26 SNPs were examined for association with CD4+ T cell count performed using PLINK for univariate and multivariate methods. The analysis used four models for testing; the allelic and three genotype tests (dominant, genotypic, and recessive). CD4+ T cell count data were available for only one-time point post-ART initiation and was not the same time point for all participants.

3.8.1 Univariate analysis of association between SNPs and CD4+ T cell count

Five SNPs were associated with CD4+ T cell count in the univariate analysis before correction for multiple testing **(Table 3.16, highlighted).** These included:

***ABCG2* rs2725252:** This is an intron variant, and it was found to be significantly associated with CD4+ T cell count in this HIV-infected SA cohort on d4T-ART, for all the four test models (allelic, genotypic, dominant and recessive; PEMP1 = **0.005**, p = **0.005**, p = **0.004**, p = **0.005**, respectively).

***SLC28A3* rs4877847:** The intron variant was found to be significantly associated with CD4+ T cell count in this HIV-infected SA cohort on d4T-ART, for all the four test models (allelic, genotypic, dominant and recessive; PEMP1 = **0.025**, p = **0.026**, p = **0.026**, p = **0.026** respectively).

SLC28A1 SNPs **rs2290272** and **rs8187758:** These missense variants of the *SLC28A1* d4T transporter gene were found to be significantly associated with CD4 count in this HIV-infected SA cohort on d4T-ART.The association of the **rs2290272** polymorphism with CD4+ T cell count occurred in all the four models tested, for the point-wise (PEMP1) p-values (allelic, genotypic, dominant and recessive; p=**0.003**, p = **0.003**, p = **0.003**, p = **0.003** respectively).

The association of the *SLC28A1* **rs8187758** polymorphism with CD4+ T cell count showed significant association for both the point-wise estimates and the family wise correction in all four models tested : PEMP1; p = **0.00163**, p = **0.00174**, p = **0.00147**, p = **0.00163**.

***TK2* rs2288399:** The thymidine kinase gene (*TK2*), dNTP pool regulator gene intron variant was found to be significantly associated with CD4+ T cell count in this HIV-infected SA cohort on d4T-ARTin all the four models tested for the point wise (PEMP1) p-values (allelic, genotypic, dominant and recessive; p=**0.027**, p=**0.027**, p=**0.028**, p=**0.027** respectively).

After correction for multiple testing, only the *SLC28A1* rs8187758 remained significantly associated with CD4+ T cell count, with PEMP2 values of 0.042 or 0.043 for all four models (Table 3.16).

Table 3.16 The univariate analysis of associations between SNPs and CD4+ T cell count.

Gene	SNP	Allelic		Genotypic		Dominant		Recessive	
		P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}
Drug transporter genes									
ABCC5	rs3749442	0.736	1.000	0.738	1.000	0.736	1.000	0.740	1.000
ABCG2	rs12505410	0.173	0.999	0.172	0.989	0.175	0.990	0.173	0.990
	rs2725252	**0.005**	0.107	**0.005**	0.105	**0.004**	0.109	**0.005**	0.106
	rs3114018	0.728	1.000	0.728	1.000	0.728	1.000	0.728	1.000
	rs2622604	0.068	0.815	0.067	0.815	0.069	0.813	0.067	0.814
SLC28A3	rs7853758	0.966	1.000	0.967	1.000	0.966	1.000	0.967	1.000
	rs4877847	**0.025**	0.475	**0.026**	0.472	**0.026**	0.472	**0.026**	0.473
SLC28A1	rs2290272	**0.003**	0.080	**0.003**	0.078	**0.003**	0.081	**0.003**	0.079
	rs8187758	**0.00163**	**0.04256**	**0.00174**	**0.0418**	**0.00147**	**0.04252**	**0.002**	**0.04295**
	rs2242046	0.250	0.999	0.253	0.999	0.249	0.999	0.253	0.999
dNTP pool regulator genes									
MTHFR	rs1801131	0.824	1.000	0.8245	1.000	0.823	1.000	0.828	1.000
	rs1801133	0.979	1.000	0.979	1.000	0.979	1.000	0.978	1.000
RRM2	rs7574663	0.271	1.000	0.271	0.999	0.268	1.00	0.266	1.000

Table 3.16 continued. The univariate analysis of associations between SNPs and CD4+ T cell count.

Gene	SNP	Allelic		Genotypic		Dominant		Recessive	
		P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}
dNTP pool regulator genes									
DHFR	rs1650723	0.099	0.920	0.097	0.918	0.097	0.918	0.099	0.918
RRM2B	rs16918482	0.454	1.000	0.4507	1.000	0.452	1.000	0.448	1.000
RRM1	rs1465952	0.755	1.000	0.755	1.000	0.754	1.000	0.758	1.000
	rs11030918	0.516	1.000	0.520	1.000	0.521	1.000	0.517	1.000
	rs12806698	0.574	1.000	0.572	1.000	0.574	1.000	0.573	1.000
	rs1042927	0.095	0.908	0.093	0.905	0.093	0.905	0.092	0.905
TK2	rs3743712	0.647	1.000	0.646	1.000	0.651	1.000	0.646	1.000
	rs11859474	0.835	1.000	0.836	1.000	0.839	1.000	0.836	1.000
	rs2288399	**0.027**	0.484	**0.027**	0.482	**0.028**	0.481	**0.027**	0.483
SAMHD1	rs8124728	0.124	0.958	0.123	0.958	0.124	0.957	0.125	0.957
	rs1291142	0.694	1.000	0.695	1.000	0.695	1.000	0.692	1.000
	rs1891643	0.847	1.000	0.847	1.000	0.848	1.000	0.846	1.000
SLC19A1	rs1051266	0.660	1.000	0.658	1.000	0.659	1.000	0.661	1.000

(Red) $p<0.05$ Significant association with CD4 count restitution.

3.8.2 Multivariate linear regression analysis of associations between SNPs and CD4+ T cell count

The multivariate linear regression analysis of associations between SNPs and CD4+ T cell count, corrected for one clinical variable (number of months HIV+). This analysis was performed using PLINK deploying four association tests: the allelic test and three genotype tests (dominant, recessive and co-dominant/genotypic). Four of the loci that were identified as significantly associated with CD4+ T cell count in the univariate analysis (*ABCG2* rs2725252, SLC28A3 rs4877847, *SLC28A1* rs2290272 and rs8187758), remained significantly associated with CD4+ T cell count in the multivariate analysis, as discussed below and Table **3.17.**However none of the SNPs remained significant in multivariate analyses after correction for multiple testing (PEMP2>0.05).

ABCG2 **rs2725252:** This polymorphism of the *ABCG2* gene was found to be significantly associated with CD4+ T cell count recovery after correcting for the number of months HIV+ in all the tests for the point-wise estimates (PEMP1) the allelic, genotypic, dominant and recessive, p=**0.019**, p=**0.011**, p=**0.078**, p=**0.015** respectively and not significantly associated with CD4 count in the family-wise correction (PEMP2) in all the four tests performed.

SLC28A3 **rs4877847:** this polymorphism of the *SLC28A3* was also found to be significantly associated with CD4+ T cell count (current) after correcting for the number of months HIV+, in three tests for the point-wise estimates (PEMP1) the allelic and genotypic, p=**0.023** and p=**0.024;** the rs4877847 polymorphism was not significantly associated with CD4 count in PEMP2 in all the four tests performed.

SLC28A1 **rs2290272 :** this polymorphism of the *SLC28A1* gene was also found to be significantly associated with CD4 count after correcting for the number of months HIV+, in all the four tests allelic, genotypic and dominant, for the point-wise estimates (PEMP1), p=**0.01**, p=**0.039** and p=**0.028** respectively, but not significantly associated with CD4 count in PEMP2.

***SLC28A1* rs8187758 :** a polymorphism of the *SLC28A1* gene was also found to found to be significantly associated with CD4 count after correcting for the number of months HIV+, in two tests for PEMP1 allelic and dominant, p=**0.002** and p=**0.029** respectively, the PEMP2 values were not significantly associated with CD4 count. This observation is different with the univariate analysis in that in the univariate analysis all the PEMP1 and PEMP2 for all the models tested were significantly associated with CD4 Count whereas with the multivariate analysis, only the allelic and the dominant tests were significantly associated with CD4 count in the point-wise estimates (PEMP1) but not significant in the family-wise correction (PEMP2).

Table 3.17 The Multivariate analysis of associations between SNPs and CD4+ T cell count.

Gene	SNP	Allelic		Genotypic		Dominant		Recessive	
		P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}
Drug transporter genes									
ABCC5	rs3749442	0.962	1.000	0.937	1.000	0.922	1.000	0.922	1.000
ABCG2	rs12505410	0.275	1.000	0.182	0.987	0.398	1.000	0.195	0.991
	rs2725252	**0.019**	0.381	**0.012**	0.223	**0.078**	0.858	**0.015**	0.305
	rs3114018	0.681	1.000	0.414	1.000	0.613	1.000	0.161	0.975
	rs2622604	0.164	0.987	0.428	1.000	0.185	0.993	0.485	1.000
SLC28A3	rs7853758	0.978	1.000	0.991	1.000	0.240	0.999	0.224	0.996
	rs4877847	**0.023**	0.441	**0.024**	0.426	0.056	0.756	0.072	0.795
SLC28A1	rs2290272	**0.012**	0.263	**0.039**	0.574	**0.028**	0.497	0.077	0.821
	rs8187758	**0.016**	0.327	0.094	0.882	**0.027**	0.479	0.151	0.970
	rs2242046	0.232	0.998	1.000	1.000	0.233	0.998	1.000	1.000
dNTP pool regulator genes									
MTHFR	rs1801131	0.965	1.000	0.787	1.000	0.971	1.000	0.778	1.000
	rs1801133	0.796	1.000	0.677	1.000	0.854	1.000	0.678	1.000
RRM2	rs7574663	0.380	1.000	0.291	0.999	0.493	1.000	0.296	1.000

Table 3.17 continued. The Multivariate analysis of associations between SNPs and CD4+ T cell count.

Gene	SNP	Allelic		Genotypic		Dominant		Recessive	
		P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}	P_{EMP1}	P_{EMP2}
dNTP pool regulator genes									
RRM2B	rs16918482	0.265	0.100	0.449	1.000	0.309	1.000	0.465	1.000
RRM1	rs1465952	0.748	1.000	0.721	1.000	0.943	1.000	0.530	1.000
	rs11030918	0.575	1.000	0.572	1.000	0.658	1.000	0.599	1.000
	rs12806698	0.245	0.998	1.000	1.000	0.246	0.999	1.000	1.000
	rs1042927	0.178	0.991	0.064	0.780	0.313	1.000	0.073	0.821
TK2	rs3743712	0.628	1.000	0.469	1.000	0.941	1.000	0.314	1.000
	rs11859474	0.988	1.000	0.886	1.000	0.9	1.000	0.849	1.000
	rs2288399	0.106	0.934	0.074	0.812	0.224	0.998	0.087	0.859
SAMHD1	rs8124728	0.077	0.860	0.323	1.000	0.085	0.879	0.412	1.000
	rs1291142	0.738	1.000	0.760	1.000	0.550	1.000	0.658	1.000
	rs1891643	0.853	1.000	0.946	1.000	0.843	1.000	0.964	1.000
SLC19A1	rs1051266	0.751	1.000	0.491	1.000	0.298	1.000	0.259	0.998

(Red) $p<0.05$ Significant association with CD4 count.

3.8.3 Analysis of associations between haplotypes and CD4+ T cell count.

Haplotype analysis was performed for all the genes with more than one SNP, using the sliding window approach in PLINK. The associations between haplotypes and CD4+ T cell count was examined in both a univariate approach, and using a multivariate linear regression approach which corrected for the number of months HIV+. Seven haplotypes in three genes were found to be associated with CD4+ T cell count recovery in uinivariate and multivariate analyses with PEMP1<0.05 before correction for multiple testing, One haplotype in ABCG2 remained significantly associated with CD4+ T cell count in univariate analysis after correction for multiple testing with PEMP2<0.05, but did not survive correction for multiple testing in the multivariate testing (**Table 3.18).**

ABCG2: A three SNP **(rs12505410|rs2725252|rs3114018**) Haplotype "TAC" of the *ABCG2* gene was found to be associated with CD4 count recovery on the univariate analysis for PEMP1 and PEMP2, p=**0.001** and p=**0.042** respectively. In the multivariate analysis this haplotype was also found to be significantly associated with CD4+ T cell count, PEMP1 = **0.004**.

The three-SNP **(rs2725252|rs3114018|rs2622604)** haplotype combinations "CAT", "ACC" and "CAC" of the *ABCG2* gene were found to be significantly associated with CD4+ T cell count recovery in the univariate analysis with PEMP1, at p=**0.009**, p=**0.002** and p=**0.009** respectively and in multivariate analysis PEMP1, at, p=**0.031**, p=**0.007** and p=**0.027** respectively.

SLC28A3: A two-SNP **(rs7853758|rs4877847)** haplotype combination "AC" of the *SLC28A3* gene was found to associated with CD4 count recovery on the univariate and multivariate analysis, for the PEMP1, p=**0.033** and p=**0.016** respectively.

SLC28A1:The three-SNP haplotype combinations "AAG" and "GCG" of the *SLC28A1* gene **(rs2290272|rs8187758Â |rs2242046)** were found to be associated with CD4 count recovery in the univariate and multivariate analysis for PEMP1, (**"AAG" p=0.003** and **p=0.002** respectively) and (**"GCG"** p=**0.002** and **p=0.008** respectively).

Table 3.18 Univariate and multivariate analysis of associations between haplotypes and CD4+ T cell count. Significant p values<0.05 shown in red.

Gene	Haplotype	SNPs in the haplotype	Univariate analysis		Multivariate analysis	
			PEMP1	PEMP2	PEMP1	PEMP2
MTHFR	GA	rs1801131|rs1801133	0.39	1.000	0.275	1.000
	TA	rs1801131|rs1801133	0.744	1.000	0.234	1.000
	GG	rs1801131|rs1801133	0.978	1.000	0.988	1.000
	TG	rs1801131|rs1801133	0.996	1.000	0.407	1.000
ABCG2	TAC	rs12505410|rs2725252|rs3114018	**0.001**	**0.042**	**0.004**	0.156
	GCC	rs12505410|rs2725252|rs3114018	0.148	0.997	0.219	1.000
	TCC	rs12505410|rs2725252|rs3114018	0.335	1.000	0.483	1.000
	TAA	rs12505410|rs2725252|rs3114018	0.475	1.000	0.697	1.000
	GCA	rs12505410|rs2725252|rs3114018	0.805	1.000	0.962	1.000
	TCA	rs12505410|rs2725252|rs3114018	0.490	1.000	0.511	1.000
	CCT	rs2725252|rs3114018|rs2622604	0.098	0.980	0.076	0.953
	CAT	rs2725252|rs3114018|rs2622604	**0.009**	0.299	**0.031**	0.700
	ACC	rs2725252|rs3114018|rs2622604	**0.002**	0.070	**0.007**	0.25
	CCC	rs2725252|rs3114018|rs2622604	0.168	0.999	0.346	1.000
	AAC	rs2725252|rs3114018|rs2622604	0.372	1.000	0.542	1.000
	CAC	rs2725252|rs3114018|rs2622604	**0.009**	0.321	**0.027**	0.653
SLC28A3	AC	rs7853758|rs4877847	**0.033**	0.725	**0.016**	0.487
	GC	rs7853758|rs4877847	0.356	1.000	0.459	1.000
	AA	rs7853758|rs4877847	0.117	0.990	0.065	0.922
	GA	rs7853758|rs4877847	0.282	1.000	0.427	1.000

Table 3.18 continued. Univariate and multivariate analysis of associations between haplotypes and CD4+ T cell count. Significant p values <0.05 shown in red.

Gene	Haplotype	SNPs in the haplotype	Univariate analysis		Multivariate analysis	
			PEMP1	PEMP2	PEMP1	PEMP2
	ACC	rs1465952|rs11030918|rs12806698	0.653	1.000	0.875	1.000
	ATC	rs1465952|rs11030918|rs12806698	0.793	1.000	0.855	1.000
	GTC	rs1465952|rs11030918|rs12806698	0.790	1.000	0.777	1.000
	CCC	rs11030918|rs12806698|rs1042927	0.320	1.000	0.473	1.000
	TCC	rs11030918|rs12806698|rs1042927	0.121	0.992	0.149	0.997
	CAA	rs11030918|rs12806698|rs1042927	0.571	1.000	0.381	1.000
	CCA	rs11030918|rs12806698|rs1042927	0.305	1.000	0.594	1.000
	TCA	rs11030918|rs12806698|rs1042927	0.644	1.000	0.642	1.000
SLC28A1	GCA	rs2290272|rs8187758Â |rs2242046	0.334	1.000	0.335	1.000
	AAG	rs2290272|rs8187758Â |rs2242046	**0.003**	0.126	**0.025**	0.632
	ACG	rs2290272|rs8187758Â |rs2242046	0.816	1.000	0.610	1.000
	GCG	rs2290272|rs8187758Â |rs2242046	**0.002**	0.082	**0.008**	0.280
TK2	GGG	rs3743712|rs11859474|rs2288399	0.128	0.994	0.275	1.000
	AGG	rs3743712|rs11859474|rs2288399	0.104	0.984	0.234	1.000
	GAA	rs3743712|rs11859474|rs2288399	0.839	1.000	0.988	1.000
	GGA	rs3743712|rs11859474|rs2288399	0.220	1.000	0.407	1.000
	AGA	rs3743712|rs11859474|rs2288399	0.253	1.000	0.326	1.000
SAMHD1	GGG	rs8124728|rs1291142|rs1891643	0.847	1.000	0.850	1.000
	GGA	rs8124728|rs1291142|rs1891643	0.596	1.000	0.664	1.000
	AAA	rs8124728|rs1291142|rs1891643	0.123	0.993	0.077	0.953
	GAA	rs8124728|rs1291142|rs1891643	0.418	1.000	0.306	1.000

(Red) p<0.05 Significant association with CD4 count restitution.

CHAPTER 4

DISCUSSION

CHAPTER 4: DISCUSSION

Antiretroviral therapy (ART) including NRTIs has changed HIV from a chronic disease to a manageable disease by curbing the death rate and the state of illness encountered by HIV+ patients (Margolis et al., 2014). However the use of ART is associated with a wide range of toxic effects such as peripheral neuropathy, lipodystrophy and pancreatitis (Kallianpur and Hulgan, 2009). Not all patients who are exposed to NRTIs develop mitochondrial toxicity such as sensory neuropathy. This suggests that factors such as underlying genetic variation can influence the development of SN in response to NRTI use.

Pharmacogenetics refers to the study of altered drug response as a result of genetic changes (Kalow, 2002; Johnson, 2003) and this may help to predict who suffers off-target side-effects of ART and who does not. In particular, human genetic variation could influence SN development via the so-called ADME genes which control **drug absorption, distribution, metabolism and excretion / elimination** (Kerb et al., 2009). These can be core determinant of the toxicities and efficacies of ARVs (Johnson, 2003; Michaud et al. 2012). In addition, many of the same ADME genes also control the relative levels of the natural dNTP pool with which the NRTIs compete inside cells. Variation in endogenous dNTP or RN levels could be a key second mechanism by NRTI toxicity could occur (Selvaraj et al., 2014).

In the current study single nucleotide polymorphisms in genes encoding d4T-transporters and enzymes regulating dTTP pools were genotyped and their association with the development of sensory neuropathy in individuals on a d4T-containing regimen was investigated. We further investigated the association of the SNPs with CD4+ T cell count and mtDNA copy number. To achieve these aims we selected 26 SNPs from the ADME genes and the SNPs were genotyped with the Sequenom assay followed by statistical associations with the phenotypes (sensory neuropathy, CD4+ T cell count, mtDNA copy number).

mtDNA copy number was investigated as an indicator of mtDNA toxicity, and a qPCR assay was used to compare mtDNA copy number in individuals with SN and without SN after d4T-use.

4.1 Genetic variation in ADME genes in the black SA population

Our reports of MAF for these SNPs in the black South African populations are novel. The *MTHFR* alleles C677T (rs1801133) and A1298C (rs1801133) have been reported in the SA populations. The *MTHFR* allele C677T (rs1801133) was reported at a frequency of 0.07 in the sub-Saharan Africa (Rosenberg et al., 2002). In this study the *MTHFR* 677T (rs1801133) polymorphism also had the same allele frequency of 0.07. Allele frequencies in our SA cohort were often but not always similar to those reported in the YRI and LWK populations. Scholtz et al. (2002) reported that the polymorphisms C677T and A1298C were less common in the black population at frequencies (0.04 and 0.09) respectively.

There were five polymorphisms in drug transporter genes that were significantly different when compared between the three populations (SA vs. LWK vs. YRI). The polymorphisms rs2725252, rs12505410 and rs3114018 of the *ABCG2* gene had relatively different minor allele frequencies compared to the YRI and LWK at p-values; $p = 0.002$, $p = 0.04$, $p = 0.04$ respectively. The rs2242046 SNP of the *SLC28A1* is monomorphic in YRI; it was polymorphic in SA with MAF of 0.025 whereas in the LWK population MAF was 0.061, hence there was a significant difference amongst the three population groups ($p = 0.0008$). The rs7853758 SNP of the *SLC28A3* with the highest frequency of 0.50 in SA was significantly different to MAF observed in the LWK and YRI ($p = 0.0001$). These differences indicate that the Nigerian or Kenyan populations are not always good proxies for South African genetic data, as others have also shown (e.g. May et al. 2013). This observation warrants the establishment of a Southern African specific genomic data set.

All the SNPs in all genes in cases and in controls were in Hardy-Weinberg equilibrium. Our data was used to construct haplotypes in seven genes and 42 haplotypes were constructed. We explored LD patterns using 21 SNPs spread across seven genes, we used confidence bounds to define a strong LD (D' is greater than **0.98**) (Gabriel et al., 2002).

4.2. Demographic factors associated with phenotypes (SN, mtDNA copy number and CD4+ T cell count after ART).

The incidence and prevalence of HIV-SN has increased in the post-ART era, suggesting that the use of ARTs, particularly NRTI use is a risk for developing sensory neuropathy (Wadley et al., 2011). Evans et al. (2011) reported that ATN is the most common toxicity of ART therapy in the sub-Saharan African. In the current study we evaluated the difference between the participants with SN and those without, in order to determine the characteristics which are independently associated with SN.

Out of the demographic and clinical variables evaluated age and height were significantly associated with SN, while the other characteristics .i.e. BMI, weight, CD4+ T cell count at sampling, CD4+ T cell count (nadir), months HIV+, were not associated with SN. We found no demographic or clinical variable that associated with mtDNA copy number. Oshinaike et al. (2012) also reported that age was independently associated with SN, before and after the use of HAART, using a Nigerian cohort attending the HIV clinic in Lagos state University teaching hospital. Several studies of Han Chinese and African populations also found age and height independently associated with SN (Cheng et al., 1999; Wadley et al., 2011; Kamerman et al., 2012). Age as a risk factor for SN is as a result of diminished activity of peripheral nerves which occurs naturally due to aging , during the stages of SN the degeneration of small and larger nerve fibres is length-dependent making taller individuals suffer the most (Oshinaike et al., 2012).

The number of months that participants had been HIV positive was significantly associated with CD4+ T cell count after ART (at sampling). CD4+ T cell count at sampling was not significantly associated with lowest CD4+ count before treatment initiation, age, height, weight, or BMI. Lawn et al. (2006) reported that advanced pre-treatment immunodeficiency characterised by a low CD4+, was associated with immune restitution even after the initiation of ART. However nadir CD4 count in our cohort was not associated with CD4+ T cell count at sampling. This could possibly be because our cohort was cross-sectional, all at different stages since ART initiation Gazzola et al. (2009) reported that old age was associated with CD4 count recovery, which is contrary to what we found in our study.

4.3 Associations between SNPs and sensory neuropathy

The univariate analysis conducted revealed that, the *SLC28A1* 709C>A rs8187758 and SAMHD1 haplotype rs8124728-rs1291142-rs1891643 GGA were associated with sensory neuropathy after ART use in a genotypic (co-dominant) model. However these associations were not significant after adjustment for multiple testing. In multivariate analyses taking into account height and age, *MTHFR* rs1801131, *RRM2B* rs16918482 and *SAMHD1* rs8124728 and rs1891643 were associated with SN after ART use, but again were not significant after correction for multiple testing. We therefore did not find any single SNP that was significantly associated with occurrence of SN after d4T use; these results were similar to GWAS results by Leger et al. (2014) who suggested that susceptibility to d4T -associated SN may not be explained by a single genetic variant with a marked effect.

SLC28A1 is a concentrative nucleoside transport which transports nucleosides and nucleoside analogues such as d4T into the epithelial cells of the renal tubule, the apical (brush-border) of the small intestines etc. (Kis et al. 2010). In the current study, the 709C>A (rs8187758) variant may have a role in SN, whereby carriers of the A allele of this SNP in co-dominance were slightly more (but not significant) susceptible to stavudine-induced neurotoxicity (PEMP1 = 0.026; OR= 0.731). rs8187758 is a missense variant (Q237K) in SLC28A1 which has been reported to increase thymidine uptake (Szöke et al., 2009). If this mutation increases d4T uptake relative to thymidine, this could underlie our observation of weak association with SN. Further functional characterization is necessary in order to have a clear understanding of the effects of the polymorphism in SN. The 709C>A (rs8187758) polymorphism was not associated with gemcitabine toxicity, a chemotherapeutic drug in a study investigating SNP association with tumour progression (Woo et al., 2012).

The *SAMHD1* gene belongs to the family of nucleases and phosphohydrolases that are distinguished by the presence of the HD domain (Gramberg et al., 2013). SAMHD1 degrades dNTPs and mediates HIV-1 restriction by decreasing the concentration of the cellular dNTP pool (Lahouassa et al. 2012).

SAMHD1 deficient myeloid cells from a patient suffering from Aicardi-Goutières syndromes have shown to be susceptible to HIV suggesting that *SAMHD1* prevents HIV infection in the myeloid cells (Coon et al., 2012). However variation in the *SAMHD1* gene was shown not to influence HIV susceptibility in Europeans and African-Americans (Coon et al. 2012). SAMHD1 dephosphorylates dTTP but not d4T (Ballana et al. 2014). Increased expression of SAMHD1 could potentially decrease the dTTP pool but not the d4T-TP pool; this relative increase in d4T-TP could perhaps increase ART efficacy and also increase ART side effects such as SN.

Decreased expression of SAMHD1 could therefore decrease ART efficacy and also decrease ART side effects such as SN. In the current study, one *SAMHD1 haplotype* (rs8124728-rs1291142-rs1891643 GGA) and two individual SNPs (rs8124728 and rs1891643) in *SAMHD1* were associated with SN in univariate and multivariate analyses respectively, but these associations were not significant after correction for multiple testing. The rs1291142 A allele is reported to decrease the mRNA levels of the SAMHD1 protein (Coon et al., 2012), whereas the functional role of the other two polymorphisms has not been determined yet.

The enzyme *MTHFR* catalyses the conversion of 5, 10 methylene-tetrahydrofolate to 5-methylene-tetrafolate and it generates methyl residues for the thymidylate synthase (TS) function (Domingo et al., 2011). The *MTHFR* 1298A→C (rs1801131) allele changes a glutamate to an alanine at position 429 of the codon (Glu429Ala) and leads to a decreased enzyme activity (Domingo et al., 2013). The activity of the MTHFR enzyme has a direct impact on the function TS (Domingo et al., 2011).

Domingo et al. (2013a) reported that the *MTHFR* gene that is devoid of polymorphisms displays a high enzyme activity which favours the formation of the 5, 10-methylene tetrahydrofolate which will consequently promote the creation of the inhibitory ternary complex involving TS, 5, 10-methylene tetrahydrofolate and deoxyuridine monophosphate. This inhibitory ternary complex will consequently lead to a decreased activity of TS which may reduce the intracellular dTTP levels and ultimately d4T-TP will have a competitive advantage (Domingo et al., 2013a). Increased d4T-TP intracellular as a result of low TS expression and high MTHFR activity is associated with SN.

In the current study this SNP was related to SN in multivariate analysis but was not significant after correction for multiple testing. Our findings were similar to those by Domingo et al. (2013a) since Domingo et al. (2013a) also found that the *MTHFR* gene polymorphisms rs1801131 was not independently associated with lipodystrophy, pancreatitis or peripheral neuropathy in Caucasian patients after d4T use.

The *MTHFR* C677T rs1801133 allele converts alanine to a valine at position 222 of the codon (Ala222Val) and is associated with the increase in the enzyme's heat-vulnerability which reduces its specific activity (Rosenberg et al., 2002). Decreased MTHFR function leads to increased function of TS and as a result the dTTP intracellular levels will be elevated and therefore d4T-TP levels will not have a competitive advantage, and subsequently reduces the risk of mtDNA toxicity or SN. We did not find an association between this SNP rs1801133 and SN. These results were also similar to those by Domingo et al. (2013a) who also found that the *MTHFR* gene polymorphism rs1801133 was not independently associated with either pancreatitis or peripheral neuropathy in Caucasian patients after d4T use. However Domingo et al. (2013b) found that combinations of *MTHFR* SNPs rs1801131 or rs1801133 with *TYMS* SNPs contributed to d4T-induced lipodystrophy in Caucasians and this therefore deserves further study in a South African context.

Ribonucleotide reductase (RNR) is the key enzyme that catalyses the *de novo* formation of deoxyribonucleotides from ribonucleotides. The RRM1 gene encodes the larger RNR1 subunit, while *RRM2B* encodes the small p53–inducible RNR2 subunit (also called p53R2). Several studies have shown that genetic variation in *RRM1* is associated with toxicity of gemcitabine, a cytosine analog used in cancer treatment (Bepler et al. 2002; Sun et al. 2006; Dong et al. 2010). Genetic variation in *RRM2B* has been shown to influence mtDNA copy number (Bourdon et al., 2007; Kimura et al., 2003; Tyynismaa et al., 2009).

In the current study the rs16918482 polymorphism of the *RRM2B* gene was related to SN in multivariate analysis but was not significant after correction for multiple testing. The function of the 3'UTR variant is unknown, but its position in the promoter region suggests that it could possible influence SN by changing RRM2B expression. This hypothesis would need to be tested.

4.4 Associations between mtDNA copy number and sensory neuropathy.

MtDNA depletion is regarded as the cornerstone of clinical NRTI toxicity, with the measurement of the mtDNA used as a surrogate to diagnose mitochondrial toxicities in tissues (Hosseini et al., 2007). DNA polymerase gamma (POLG/*Pol-γ*) is part of the enzymes and proteins responsible for mtDNA replication and is inhibited by NRTIs (Lewis et al., 2003; Selvaraj et al., 2014).It has been demonstrated that the depletion of the mtDNA is as a result of the inhibition of *POLG* by the NRTIs, however studies have not been able to categorically establish the link between NRTIs and the loss of mtDNA content, which implies that other secondary factors also play an effective role of causing SN (Selvaraj et al., 2014). The mtDNA/nDNA ratios of cases (with SN) were compared with those of controls (without SN).

Dalakas et al. (2001) showed that mtDNA copy number was depleted in nerve cells during ART- induced SN. While the gold standard for the diagnosis of ART - related mitochondrial toxic effects is muscle, fat or liver biopsy, this is not practical for routine screening and monitoring (Côté et al., 2002) and the detection of mtDNA changes in blood samples would be preferable. Kampira et al. (2014) found significantly lowered mtDNA copy number in whole blood samples in association with ART- induced SN and lactatemia. Cherry et al. (2002) showed that ART-induced lactatemia was associated with decreased mtDNA copy number in adipose tissue but not in PBMCs.

In the current study, there were no significant differences in mtDNA copy number in whole blood samples from South African patients with or without SN after d4T use. The diagnosis of SN in this study differed from the one in Kampira et al. (2014), which might be the reason for the contradicting results. Results obtained in the current study were similar to those by Cherry et al. (2002). Urata et al. (2008) observed that mtDNA copy number in PBMCs varied due to platelet contamination and this caused overestimation of mtDNA content. Therefore, measurement of mtDNA copy number in whole blood samples which includes platelets (as done in this study), may not be very accurate.

4.5 Associations between SNPs and mtDNA copy number

All the polymorphisms in the d4T transporter genes and dNTP pool regulator genes were not significantly associated with mtDNA copy number, in the univariate analysis and Taken together with the lack of variation of mtDNA copy number in cases and controls described above, this indicates that these SNPs were not mediating SN via the mechanism of mitochondrial toxicity.

4.6 Associations between SNPs and CD4 + T cell count

There is substantial amount of inter-individual differences observed in terms of CD4+ T cell count restitution after initiating ART (Grady et al., 2011). Variation in mtDNA, HLA, IL10 and IL7R genes has been shown to influence CD4+ T cell recovery after ART (Grady et al., 2011; Kampira et al., 2013; Kampira et al., 2014; Naicker et al., 2015; Haralambieva et al., 2014; Chopera et al., 2017). The influence of variation in drug transporter *ABCB1* (*MDR1*) on CD4+ T cell count after ART has been inconsistent, with Fellay et al. (2002) and Parathyras et al. (2009) showing that variants in *ABCB1* associated with immunological recovery after ART, whereas Nasi et al. (2003) and Winzer et al. (2005) did not find similar results.

Few other ADME genes have been examined in the context of CD4+ T cell count recovery after ART. We investigated the effect of variation in genes encoding d4T transporters, and genes influencing thymidine metabolism, on CD4+ T cell count after d4T-ART in this cohort.

During univariate analysis, five SNPs and seven haplotypes in the *ABCG2*, *SLC28A1*, *SLC28A3* and *TK2* genes were associated with CD4+ T cell count after ART use. The *SLC28A1* SNP rs81877158 and one haplotype in *ABCG2* (rs12505410-rs2725252-rs3114018 TAC) remained significantly associated with CD4+ T cell count after correction for multiple testing. In multivariate analyses taking into account months since HIV diagnosis, four SNPs and the same seven haplotypes were associated with CD4+ T cell count after ART use, but were not significant after correction for multiple testing.

The *SLC28A1* SNP rs81877158 was discussed in the previous section. We found that it was associated with increased CD4+ T cell count. It was the only SNP in the current study that survived corrections for multiple testing and had potential roles in both SN and CD4+ T cell counts after ART use. This SNP was not associated with changes in mtDNA copy number and therefore would not mediate CD4 / SN phenotypes by the mechanism of mitochondrial toxicity, As mentioned, rs8187758 is a missense variant (Q237K) which has been reported to increase thymidine uptake of the SLC28A1 transporter protein (Szöke et al., 2009). It is not known if d4T uptake is also changed by this mutation. Elevated intracellular levels of d4T-TP will reinforce the ability of antiviral drug to inhibit HIV-1 viral replication and as a result diminish HIV-1 viral the improved drug efficacy will consequently enable the increase of CD4+ T cells and therefore immune restoration. Further functional characterization is necessary in order to have a clear understanding of the effects of the polymorphism on CD4+ T cell count changes.

The BCRP transporter, encoded by the *ABCG2* gene, is widely expressed in the liver, gastrointestinal tract and is involved in the efflux of clinically used drugs (Delord et al., 2013). This efflux transporter of nucleosides may have a protective role against adverse drug events such as mtDNA toxicity by eliminating toxic substances and metabolites across the apical membrane and preventing their intracellular accumulation (Maekawa et al., 2006). Increased efflux activity could cause decreased intracellular ARV amounts and therefore decreased ART efficacy (e.g. lower CD4 count), or decreased ART-induced toxicity, whereas decreased efflux activity could cause increased intracellular ARV amounts and therefore increased ART efficacy or increased toxicity.

In this study the rs12505410|rs2725252|rs3114018 TAC haplotype in *ABCG2* was associated with increased CD4+ T cell count and this remained significant after correction for multiple testing. Three other haplotypes in ABCG2 (rs2725252|rs3114018|rs2622604 CAT, ACC and CAC) were associated with increased CD4+ T cell count but did not remain significant after correction for multiple testing. In addition, the single *ABCG2* SNP rs2725252 was associated with increased CD4+ T cell count in univariate and multivariate analysis (PEMP<0.05), but this was not significant after correction for multiple testing.

The functions of the *ABCG2* SNPs rs12505410, rs2725252, and rs3114018 are associated with changes in mRNA expression, the functionality of these SNPs is sparsely reported on and therefore it remains ambiguous as to whether the changes in mRNA expression are elevated or reduced. In addition, rs3114018 (C to A) has been associated with oxaliplatin-induced peripheral neuropathy in the Caucasian population of Spain (Custodio et al., 2014), suggesting a potential role for this SNP in decreased drug efflux. The rs12505410 - rs2725252 GG haplotype has been associated with response to imatinib during treatment of chronic myelogenous leukaemia (Delord et al. 2013).

Delord et al. (2010) discovered that the *ABCG2*,19-744G>T (rs2725252) polymorphism was associated with the response of the 400 mg/d arm imatinib, a drug used for the treatment of chronic myelogenous leukaemia (CLM) together with two other *ABCG2* polymorphisms (rs12505410 and rs13120400) in a Saint Louis exploratory cohort, Paris. We hypothesize that the *ABCG2* haplotype rs12505410|rs2725252|rs3114018 TAC is associated with decreased ABCG2 function, and therefore decreased ARV efflux, resulting in increased intracellular levels of d4T-TP and consequently good CD4+ count response. This hypothesis would need to be tested using functional assays.

Transporter protein CNT3 encoded by the *SLC28A3* gene facilitates d4T influx into cells (Molina-Arcas, Casado and Pastor-Anglada, 2009). In this study the rs4877847 SNP was associated with decreased CD4+ T cell count in both univariate and multivariate analyses but was not significant after correction for multiple testing. In addition one haplotype (rs7853758-rs4877847 AC) was associated with decreased CD4+ T cell count but was not significant after correction for multiple testing.

The rs4877847 SNP has been reported to cause loss of function of the influx transporter (Visscher et al. 2011). The 1502A>C (rs4877847) polymorphism was also revealed to be associated with a protective effect in anthracycline-induced cardiotoxicity in a cohort of Canadian descent (Visscher et al., 2013). If this SNP causes decreased *SLC28A3*-mediated influx of d4T, this could potentially cause

decreased intracellular concentrations of ARV and therefore decreased CD4 counts. This hypothesis would need to be tested using functional assays.

The *TK2* gene is expressed ubiquitously in mitochondria and is responsible for the phosphorylation of deoxythymidine and deoxycytidine contributing in the salvage pathway of deoxynucleotide synthesis in the mitochondria (Oskoui et al., 2006). The initial phosphorylation of NRTIs is very important for both their pharmacological functional and mitochondrial toxicity (Hosseini et al., 2007). Mutations in *TK2* have been associated with autosomal recessive mitochondrial DNA depletion syndrome-2 (MTDPS2), mitochondrial myopathy and progressive external ophthalmoplegia (PEO) (e.g. Saada et al., 2001; Behin et al., 2012).

In this study the rs2288399 SNP was associated with decreased CD4+ T cell count in univariate analyses but was not significant after correction for multiple testing. This SNP is an intron variant with unknown function and few studies have previously examined this SNP. It was chosen in this study because it tags many (23) other SNPs in *TK2* and was present in 15% of the SA cohort based on our results.

4.7. Strengths and weaknesses of the study

4.7.1 Strengths

The study examined polymorphisms that have never been studied before in the Southern African population, it has revealed very interesting findings one of them being that, the Southern African population is different to the other African populations on certain SNPs and a genomic data set specific for the southern African population is necessary. Few studies looked at ADME genetics underlying CD4+ T cell response, which is critical because the sub-Saharan Africa has the highest number of people living with HIV and on ART.

4.7.2 Weaknesses

This study had the following weaknesses worth noting. The clinical information presented was only retrospective for a single time-point and no prospective clinical data was attained with respect to measures such as CD4+ count, viral load, treatment ability and peripheral pain. The phenotypic data available were extremely limited, and did not include other measures of mitochondrial toxicity such as lactate levels or measures of lipodystrophy. SN cases and controls were not matched with regards to age and height, but these factors were taken into account in multivariate analysis.

The DNA was the only sample we had; we did not have other samples like the plasma to check the dug levels of ART or cells for measuring gene expression. Samples of various types should be collected in future in order to perform functional assays. The YRI and LWK are not always good proxies for genetic data in SA populations and therefore reliance of these populations as a reference for the SA population does not always good results.

This study examined only 26 SNPs; whole gene sequences or whole genome SNP approaches would be preferable. Gene- gene and gene-environment interactions could also be analyzed, Due to the limitations of the current study, a replication study investigating the same polymorphisms can be carried out in a similar ART treated patient cohort in SA and other populations to confirm and validate our findings.

The replacement of d4T by tenofovir in current SA ART regimens may mean that the current findings may no longer be relevant to clinical practice.

Conclusion

Identifying predictors of response to ART in South Africa is a priority since > 3 million people are using ART in this country in 2016. In particular, human genetic variation could influence SN development or CD4+ T cell recovery after ART via the ADME genes which control drug absorption, distribution, metabolism and excretion / elimination. Our study took a candidate gene system approach which investigated variation in genes that transport stavudine and genes that modulate the levels of dNTPs in the cell, and examining their association with SN, with mitochondrial toxicity as measured by mtDNA:nDNA ratio, and with CD4+ T cell count after ART initiation in a South African cohort using d4T-ART.

We determined allele frequencies for 26 SNPs in seven ADME genes that could form a basis for future studies in pharmacogenetic applications in South African populations. We showed that allele frequencies in South Africans are not always the same as those in the often-used proxy Kenya and Nigerian populations.

We confirmed that in this cohort, age and height were significantly associated with SN after d4T use, and that time since HIV diagnosis / ART initiation was significantly associated with post-ART CD4 count. This study identified no significant associations between 26 SNPs and SN, or between the 26 SNPs and mtDNA copy number. We also found no association between mtDNA copy measured in DNA extracted from whole blood, and SN.

We found that the *SLC28A1* SNP rs8187758 and an *ABCG2* haplotype were significantly associated with the CD4+ T cell count after stavudine use. This finding proposes a potential predictor of ART drug response in South Africans which may be used to help formulate drug doses that will be optimal for personalized treatment. These SNPs should be examined in South Africans using more current (tenofovir-based) ART regimens to determine whether testing for these SNPs may still be useful.

REFERENCES

Agena bioscience 2015, iPLEX genotyping chemistry accessed May 7 2015<http://agenabio.com/pharmacogenetics>.

Aiuti, F. and Mezzaroma, I., 2006. Failure to reconstitute CD4+ T-cells despite suppression of HIV replication under HAART. *AIDS* Rev, **8**(2), pp.88-97.

Amie, S.M., Daly, M.B., Noble, E., Schinazi, R.F., Bambara, R.A. and Kim, B., 2013. Anti-HIV host factor SAMHD1 regulates viral sensitivity to nucleoside reverse transcriptase inhibitors via modulation of cellular deoxyribonucleoside triphosphate (dNTP) levels. *The Journal of Biological Chemistry,* **288**(28), pp. 20683-20691.

Andreu, A.L., Martinez, R., Marti, R. and García-Arumí, E., 2009. Quantification of mitochondrial DNA copy number: pre-analytical factors. *Mitochondrion,* **9**(4), pp. 242-246.

Arab-Alameddine, M., Décosterd, L.A., Buclin, T., Telenti, A. and Csajka, C., 2011. Antiretroviral drug toxicity in relation to pharmacokinetics, metabolic profile and pharmacogenetics. *Expert Opinion on Drug Metabolism & Toxicology*, **7**(5), pp.609-622.

Arts, E.J. and Wainberg, M.A., 1996. Mechanisms of nucleoside analog antiviral activity and resistance during human immunodeficiency virus reverse transcription. *Antimicrobial Agents and Chemotherapy,* **40**(3), pp. 527-540.

Awofala, A.A. and Ogundele, O.E., 2016. HIV epidemiology in Nigeria. Saudi *Journal of Biological Sciences,* http://dx.doi.org/10.1016/j.sjbs.2016.03.00 August 21, 2017.

Back, D., 1999. Pharmacological issues relating to viral resistance. *Infection,* **27**(2), pp. S42-S44.

Badagnani, I., Chan, W., Castro, R., Brett, C., Huang, C., STRYKE, D., Kawamoto, M., Johns, S., Ferrin, T. and Carlson, E., 2005. Functional analysis of genetic variants in the human concentrative nucleoside transporter 3 (CNT3; SLC28A3). *The Pharmacogenomics Journal,* **5**(3), pp. 157-165.

Bailey, L.B. ed., 2009. *Folate in Health and Disease*. CRC Press.

Baker, J.V., Peng, G., Rapkin, J., Abrams, D.I., Silverberg, M.J., MacArthur, R.D., Cavert, W.P., Henry, W.K. and Neaton, J.D., 2008. CD4+ count and risk of non-AIDS diseases following initial treatment for HIV infection. *AIDS* (London, England), 22(7), p.841.

Ballana, E., Badia, R., Terradas, G., Torres-Torronteras, J., Ruiz, A., Pauls, E., et al., 2014. SAMHD1 specifically affects the antiviral potency of thymidine analog HIV reverse transcriptase inhibitors. *Antimicrobial Agents and Chemotherapy,* **58**(8), pp. 4804-4813.

Barrett, J.C., Fry, B., Maller, J.D.M.J. and Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), pp.263-265.

Barrett, J.C., Fry, B., Maller, J.D.M.J. and Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), pp.263-265.

Behin, A., Jardel, C., Claeys, K.G., Fagart, J., Louha, M., Romero, N.B., Laforet, P., Eymard, B. and Lombes, A., 2012. Adult cases of mitochondrial DNA depletion due to TK2 defect an expanding spectrum. *Neurology*, **78**(9), pp.644-648.

Bonod-Bidaud, C., Chevrollier, A., Bourasseau, I., Lachaux, A., de Camaret, B.M. and Stepien, G., 2001. Induction of ANT2 gene expression in liver of patients with mitochondrial DNA depletion. *Mitochondrion*, *1*(3), pp.217-224.

Borst, P. and Elferink, R.O., 2002. Mammalian ABC transporters in health and disease. *Annual Review of Biochemistry,* **71**(1), pp. 537-592.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics,* **32**(3), pp. 314-331.

Bourdon, A., Minai, L., Serre, V., Jais, J.P., Sarzi, E., Aubert, S., Chrétien, D., de Lonlay, P., Paquis-Flucklinger, V., Arakawa, H. and Nakamura, Y., 2007. Mutation of

RRM2B, encoding p53-controlled ribonucleotide reductase (p53R2), causes severe mitochondrial DNA depletion. *Nature Genetics*, 39(6), pp.776-780.

Brinkman, K., Smeitink, J.A., Romijn, J.A. and Reiss, P., 1999. Mitochondrial toxicity induced by nucleoside-analogue reverse-transcriptase inhibitors is a key factor in the pathogenesis of antiretroviral-therapy-related lipodystrophy. *The Lancet,* **354**(9184), pp. 1112-1115.

Burckhardt, G., 2012. Drug transport by organic anion transporters (OATs). *Pharmacology & Therapeutics,* **136**(1), pp. 106-130.

Calza, L., Manfredi, R. and Chiodo, F., 2005. Hyperlactataemia and lactic acidosis in HIV-infected patients receiving antiretroviral therapy. *Clinical Nutrition*, *24*(1), pp.5-15.

Çam, R., Eroglu, A., Egin, Y. and Akar, N., 2009. Dihydrofolate reductase (DHRF) 19-bp intron-1 deletion and methylenetetrahydrofolate reductase (MTHFR) C677T polymorphisms in breast cancer. *Breast Cancer Research and Treatment*, *115*(2), pp.431-432.

Cano-soldado, P. and Pastor-anglada, M., 2012. Transporters that translocate nucleosides and structural similar drugs: structural requirements for substrate recognition. *Medicinal Research Reviews,* **32**(2), pp. 428-457.

Cano-Soldado, P., Larráyoz, I.M., Molina-Arcas, M., Casado, F.J., Martinez-Picado, J., Lostao, M.P. and Pastor-AngladA, M., 2004. Interaction of nucleoside inhibitors of HIV-1 reverse transcriptase with the concentrative nucleoside transporter-1 (SLC28A1). *Antiviral Therapy*, **9**(6), pp.993-1002.

Canter, J.A., Haas, D.W., Kallianpur, A.R., Ritchie, M.D., Robbins, G.K., Shafer, R.W., Clifford, D.B., Murdock, D.G. and Hulgan, T., 2008. The mitochondrial pharmacogenomics of haplogroup T: MTND2* LHON4917G and antiretroviral therapy-associated peripheral neuropathy. *The Pharmacogenomics Journal*, 8(1), pp.71-77.

Cao, X., Mitra, A.K., Pounds, S., Crews, K.R., Gandhi, V., Plunkett, W., Dolan, M.E., Hartford, C., Raimondi, S., Campana, D. and Downing, J., 2013. RRM1 and RRM2

pharmacogenetics: association with phenotypes in HapMap cell lines and acute myeloid leukemia patients. *Pharmacogenomics*, 14(12), pp.1449-1466.

Carr, A. and Cooper, D.A., 2000. Adverse effects of antiretroviral therapy. *The Lancet,* **356**(9239), pp. 1423-1430.

Cascorbi, I., 2006. Role of pharmacogenetics of ATP-binding cassette transporters in the pharmacokinetics of drugs. *Pharmacology & Therapeutics,* **112**(2), pp. 457-473.

Cha, P.C., Mushiroda, T., Zembutsu, H., Harada, H., Shinoda, N., Kawamoto, S., Shimoyama, R., Nishidate, T., Furuhata, T., Sasaki, K. and Hirata, K., 2009. Single nucleotide polymorphism in *ABCG2* is associated with irinotecan-induced severe myelosuppression. *Journal of Human Genetics*, 54(10), pp.572-580.

Chan, L.M., Lowes, S. and Hirst, B.H., 2004. The ABCs of drug transport in intestine and liver: efflux proteins limiting drug absorption and bioavailability. *European Journal of Pharmaceutical Sciences*, 21(1), pp.25-51.

Cherry, C.L., Gahan, M.E., McArthur, J.C., Lewin, S.R., Hoy, J.F. and Wesselingh, S.L., 2002. Exposure to dideoxynucleosides is reflected in lowered mitochondrial DNA in subcutaneous fat. *JAIDS-HAGERSTOWN MD-,* 30(3), pp.271-277.

Cherry, C.L., Mcarthur, J.C., Hoy, J.F. and Wesselingh, S.L., 2003. Nucleoside analogues and neuropathy in the era of HAART. *Journal of Clinical Virology,* **26**(2), pp. 195-207.

Cherry, C.L., Rosenow, A., Affandi, J.S., McArthur, J.C., Wesselingh, S.L. and Price, P., 2008. Cytokine genotype suggests a role for inflammation in nucleoside analog-associated sensory neuropathy (NRTI-SN) and predicts an individual's NRTI-SN risk. *AIDS Research and Human Retroviruses,* **24**(2), pp. 117-123.

Chew, C., Cherry, C., Imran, D., Yunihastuti, E., Kamarulzaman, A., Varna, S., Ismail, R., Phipps, M., Aghafar, Z. and Gut, I., 2011. Tumour necrosis factor haplotypes associated with sensory neuropathy in Asian and Caucasian human immunodeficiency virus patients. *Tissue Antigens,* **77**(2), pp. 126-130.

Chopera, D.R., Ntale, R., Ndabambi, N., Garrett, N., Gray, C.M., Matten, D., Karim, Q.A., Karim, S.A. and Williamson, C., 2017. Early evolution of human leucocyte antigen-associated escape mutations in variable Gag proteins predicts CD4+ decline in HIV-1 subtype C-infected women. *AIDS*, **31**(2), pp.191-197.

Choudhuri, S. and Klaassen, C.D., 2006. Structure, function, expression, genomic organization, and single nucleotide polymorphisms of human ABCB1 (MDR1), ABCC (MRP), and ABCG2 (BCRP) efflux transporters. *International Journal of Toxicology,* **25**(4), pp. 231-259.

Christensen, E.R., Stegger, M., Jensen-Fangel, S., Laursen, A.L. and Ostergaard, L., 2004. Mitochondrial DNA levels in fat and blood cells from patients with lipodystrophy or peripheral neuropathy and the effect of 90 days of high-dose coenzyme Q treatment: a randomized, double-blind, placebo-controlled pilot study. *Clinical Infectious Diseases*, **39**(9), pp.1371-1379.

Christiansen, L.S., Munch-petersen, B. and Knecht, W., 2015. Non-Viral Deoxyribonucleoside Kinases - Diversity and Practical Use. *Journal of Genetics and Genomics*, doi:10.1016/j.jgg.2015.01.003.

Coon, S., Wang, D. and Wu, L., 2012. Polymorphisms of the SAMHD1 gene are not associated with the infection and natural control of HIV type 1 in Europeans and African-Americans. *AIDS Research and Human Retroviruses*, **28**(12), pp.1565-1573.

Cooray, H.C., Blackmore, C.G., Maskell, L. and Barrand, M.A., 2002. Localisation of breast cancer resistance protein in microvessel endothelium of human brain. *Neuroreport*, **13**(16), pp.2059-2063.

Cortright, R.N., Zheng, D., Jones, J.P., Fluckey, J.D., Dicarlo, S.E., Grujic, D., Lowell, B.B. and Dohm, G.L., 1999. Regulation of skeletal muscle UCP-2 and UCP-3 gene expression by exercise and denervation. *The American Journal of Physiology,* **276**(1 Pt 1), pp. E217-21.

Côté, H.C., Brumme, Z.L., Craib, K.J., Alexander, C.S., Wynhoven, B., Ting, L., Wong, H., Harris, M., Harrigan, P.R., O'shaughnessy, M.V. and Montaner, J.S., 2002.

Changes in mitochondrial DNA as a marker of nucleoside toxicity in HIV-infected patients. *New England Journal of Medicine*, **346(**11), pp.811-820.

Dalakas, M.C., 2001. Peripheral neuropathy and antiretroviral drugs. *Journal of the Peripheral Nervous System,* **6**(1), pp. 14-20.

Dalakas, M.C., Semino-Mora, C. and Leon-Monzon, M., 2001. Mitochondrial alterations with mitochondrial DNA depletion in the nerves of AIDS patients with peripheral neuropathy induced by 2′ 3′-dideoxycytidine (ddC). *Laboratory Investigation*, **81**(11), pp.1537-1544.

Dandara, C., Lombard, Z., Du Plooy, I., McLellan, T., Norris, S.A. and Ramsay, M., 2011. Genetic variants in CYP (-1A2,-2C9,-2C19,-3A4 and-3A5), VKORC1 and ABCB1 genes in a black South African population: a window into diversity. *Pharmacogenomics*, **12**(12), pp.1663-1670.

de Béthune, M., 2010. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989–2009). *Antiviral Research,* **85**(1), pp. 75-90.

De Cock, K.M., Jaffe, H.W. and Curran, J.W., 2012. The evolving epidemiology of HIV/AIDS. *AIDS (London, England),* **26**(10), pp. 1205-1213.

De Gorter, M., Xia, C., Yang, J. and Kim, R., 2012. Drug transporters in drug efficacy and toxicity. *Annual Review of Pharmacology and Toxicology,* **52**, pp. 249-273.

de Silva, N. and Davis, B., 2013. Iron, B12 and folate. *Medicine,* **41**(4), pp. 204-207.

Delord, M., Rousselot, P., Cayuela, J.M., Sigaux, F., Guilhot, J., Preudhomme, C., Guilhot, F., Loiseau, P., Raffoux, E., Geromin, D. and Génin, E., 2013. High imatinib dose overcomes insufficient response associated with ABCG2 haplotype in chronic myelogenous leukemia patients. *Oncotarget*, **4**(10), p.1582.

Di Martino, M.T., Arbitrio, M., Leone, E., Guzzi, P.H., Saveria Rotundo, M., Ciliberto, D., Tomaino, V., Fabiani, F., Talarico, D. and Sperlongano, P., 2011. Single nucleotide polymorphisms of ABCC5 and ABCG1 transporter genes correlate to irinotecan-

associated gastrointestinal toxicity in colorectal cancer patients: a DMET microarray profiling study. *Cancer Biology & Therapy,* **12**(9), pp. 780-787.

Domingo, P., Cabeza, M.C., Pruvost, A., Torres, F., Salazar, J., Del mar Gutierrez, M., et al., 2011. Association of thymidylate synthase gene polymorphisms with stavudine triphosphate intracellular levels and lipodystrophy. *Antimicrobial Agents and Chemotherapy,* **55**(4), pp. 1428-1435.

Domingo, P., del Carmen Cabeza, M., Torres, F., Salazar, J., del Mar Gutierrez, M., Mateo, M.G., Martínez, E., Domingo, J.C., Fernandez, I., Villarroya, F. and Ribera, E., 2013a. Association of thymidylate synthase polymorphisms with acute pancreatitis and/or peripheral neuropathy in HIV-infected patients on stavudine-based therapy. *PloS one*, *8*(2), p.e57347.

Domingo, P., Mateo, M.G., Pruvost, A., Torres, F., Salazar, J., del Mar Gutierrez, M., et al., 2013b. Polymorphisms of Pyrimidine Pathway Enzymes Encoding Genes and HLA-B* 40：01 Carriage in Stavudine-Associated Lipodystrophy in HIV-Infected Patients. *PloS One,* **8**(6), pp. e67035.

Dudley, M.N., Graham, K.K., Kaul, S., Geletko, S., Dunkle, L. and Mayer, K., 1992. Pharmacokinetics of stavudine in patients with AIDS or AIDS-related complex. *Journal of Infectious Diseases*, *166*(3), pp.480-485.

Elens, L., Vandercam, B., Yombi, J.C., Lison, D., Wallemacq, P. and Haufroid, V., 2010. Influence of host genetic factors on efavirenz plasma and intracellular pharmacokinetics in HIV-1-infected patients. *Pharmacogenomics*, 11(9), pp.1223-1234.

El-Hattab, A.W. and Scaglia, F., 2013. Mitochondrial DNA depletion syndromes: review and updates of genetic basis, manifestations, and therapeutic options. *Neurotherapeutics*, 10(2), pp.186-198.

Estudante, M., Morais, J.G., Soveral, G. and Benet, L.Z., 2013. Intestinal drug transporters: an overview. *Advanced Drug Delivery Reviews,* **65**(10), pp. 1340-1356.

Fellay, J., Marzolini, C., Meaden, E.R., Back, D.J., Buclin, T., Chave, J.P., Decosterd, L.A., Furrer, H., Opravil, M., Pantaleo, G. and Retelska, D., 2002. Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study. *The Lancet*, **359**(9300), pp.30-36.

Fukuda, Y. and Schuetz, J.D., 2012. ABC transporters and their role in nucleoside and nucleotide drug resistance. *Biochemical Pharmacology,* **83**(8), pp. 1073-1083.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. and Liu-Cordero, S.N., 2002. The structure of haplotype blocks in the human genome. *Science*, 296(5576), pp.2225-2229.

Gallant, J.E., 2002. Initial therapy of HIV infection. *Journal of Clinical Virology,* **25**(3), pp. 317-333.

Gamazon, E.R., Ziliak, D., Im, H.K., LaCroix, B., Park, D.S., Cox, N.J. and Huang, R.S., 2012. Genetic architecture of microRNA expression: implications for the transcriptome and complex traits. *The American Journal of Human Genetics*, *90*(6), pp.1046-1063.

Garg, M., Dutta, T. and Jain, N.K., 2007. Reduced hepatic toxicity enhanced cellular uptake and altered pharmacokinetics of stavudine loaded galactosylated liposomes. *European Journal of Pharmaceutics and Biopharmaceutics,* **67**(1), pp. 76-85.

Genetics Home Reference 2015, dihydrofolate reductase; DHFR, accessed May 7, 2015<http://ghr.nlm.nih.gov/gene/DHFR>.

Genetics Home Reference 2015, PINK1, accessed May 7, 2015<http://ghr.nlm.nih.gov/gene/PINK1>.

Gonzalez-Duarte, A., CikureL, K. and Simpson, D.M., 2007. Managing HIV peripheral neuropathy. *Current HIV/AIDS Reports,* **4**(3), pp. 114-118.

Goole, J., Lindley, D.J., Roth, W., Carl, S.M., Amighi, K., Kauffmann, J. and Knipp, G.T., 2010. The effects of excipients on transporter mediated absorption. *International Journal of Pharmaceutics,* **393**(1), pp. 17-31.

Goyette, P., Pai, A., Milos, R., Frosst, P., Tran, P., Chen, Z., Chan, M. and Rozen, R., 1998. Gene structure of human and mouse methylenetetrahydrofolate reductase (MTHFR). *Mammalian Genome,* **9**(8), pp. 652-656.

Grady, B.J., Samuels, D.C., Robbins, G.K., Selph, D., Canter, J.A., Pollard, R.B., Haas, D.W., Shafer, R., Kalams, S.A., Murdock, D.G. and Ritchie, M.D., 2011. Mitochondrial genomics and CD4 T-cell count recovery after antiretroviral therapy initiation in AIDS Clinical Trials Group Study 384. *Journal of Acquired Immune Deficiency Syndromes* (1999), 58(4), p.363.

Guittet, O., Hakansson, P., Voevodskaya, N., Fridd, S., Graslund, A., Arakawa, H., et al., 2001. Mammalian p53R2 protein forms an active ribonucleotide reductase in vitro with the R1 protein, which is expressed both in resting cells in response to DNA damage and in proliferating cells. *The Journal of Biological Chemistry,* **276**(44), pp. 40647-40651.

Hampras, S.S., Sucheston, L., Weiss, J., Baer, M.R., Zirpoli, G., Singh, P.K., Wetzler, M., Chennamaneni, R., Blanco, J.G., Ford, L. and Moysich, K.B., 2010. Genetic polymorphisms of ATP-binding cassette (ABC) proteins, overall survival and drug toxicity in patients with Acute Myeloid Leukemia. *International Journal of Molecular Epidemiology and Genetics*, *1*(3), p.201.

Hao, S., 2013. The Molecular and Pharmacological Mechanisms of HIV-Related Neuropathic Pain. *Current Neuropharmacology,* **11**(5), pp. 499-512.

Haralambieva, I.H., Lambert, N.D., Ovsyannikova, I.G., Kennedy, R.B., Larrabee, B.R., Pankratz, V.S. and Poland, G.A., 2014. Associations between single nucleotide polymorphisms in cellular viral receptors and attachment factor-related genes and humoral immunity to rubella vaccination. *PloS one*, **9**(6), p.e99997.

Hauschka, P.V., 1973. Analysis of nucleotide pools in animal cells. *Methods Cell Biology,* **7**, pp. 361-462.

Hediger, M.A., Clémençon, B., Burrier, R.E. and Bruford, E.A., 2013. The ABCs of membrane transporters in health and disease (SLC series): Introduction. *Molecular Aspects of Medicine,* **34**(2–3), pp. 95-107.

Hediger, M.A., Romero, M.F., Peng, J., Rolfs, A., Takanaga, H. and Bruford, E.A., 2004. The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflügers Archive,* **447**(5), pp. 465-468.

Hendry, L., Lombard, Z., Wadley, A. and Kamerman, P., 2013. KCNS1, but not GCH1, is associated with pain intensity in a black southern African population with HIV-associated sensory neuropathy: a genetic association study. *Journal of Acquired Immune Deficiency Syndromes (1999),* **63**(1), pp. 27-30.

Hendry, L.M., 2014. *Genetics of HIV-associated sensory neuropathy in black Southern Africans,* (Doctoral dissertation), University of the Witwatersrand.

Ho, V., Massey, T.E. and King, W.D., 2011. Thymidylate synthase gene polymorphisms and markers of DNA methylation capacity. *Molecular Genetics and Metabolism,* **102**(4), pp. 481-487.

Hurst, M. and Noble, S., 1999. Stavudine. *Drugs*, *58*(5), pp.919-949.

Imamichi, T., 2004. Action of anti-HIV drugs and resistance: reverse transcriptase inhibitors and protease inhibitors. *Current Pharmaceutical Design,* **10**(32), pp. 4039-4053.

Ishikawa, T., Tsuji, A., Inui, K., Sai, Y., Anzai, N., Wada, M., Endou, H. and Sumino, Y., 2004. The genetic polymorphism of drug transporters: functional analysis approaches. *Pharmacogenomics*, *5*(1), pp.67-99.

Johnson, J.A., 2003. Pharmacogenetics: potential for individualized drug therapy through genetics. *TRENDS in Genetics,* **19**(11), pp. 660-666.

Johnson, L.F., 2012. Access to antiretroviral treatment in South Africa, 2004-2011. *Southern African Journal of HIV Medicine,* **13**(1).

Joint United Nations Programme on HIV and AIDS, Joint United Nations on HIV and AIDS Factsheet 2015: Word AIDS Day 2015.

Kalow, W., 2002. Pharmacogenetics and personalised medicine. *Fundamental & Clinical Pharmacology,* **16**(5), pp. 337-342.

Kamerman, P.R., Wadley, A.L. and Cherry, C.L., 2012. HIV-associated sensory neuropathy: risk factors and genetics. *Current Pain and Headache Reports,* **16**(3), pp. 226-236.

Kampira, E., Dzobo, K., Kumwenda, J., van Oosterhout, J.J., Parker, M.I. and Dandara, C., 2014. Peripheral blood mitochondrial DNA/nuclear DNA (mtDNA/nDNA) ratio as a marker of mitochondrial toxicities of stavudine containing antiretroviral therapy in HIV-infected Malawian patients. *Omics: a Journal of Integrative Biology*, 18(7), pp.438-445.

Kampira, E., Kumwenda, J., van Oosterhout, J.J. and Dandara, C., 2013. Mitochondrial DNA subhaplogroups L0a2 and L2a modify susceptibility to peripheral neuropathy in Malawian adults on stavudine containing highly active antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes* (1999), 63(5), p.647.

Kerb, R., Fux, R., Mörike, K., Kremsner, P.G., Gil, J.P., Gleiter, C.H. and Schwab, M., 2009. Pharmacogenetics of antimalarial drugs: effect on metabolism and transport. *The Lancet Infectious Diseases*, 9(12), pp.760-774.

Keswani, S.C., Pardo, C.A., Cherry, C.L., HOKE, A. and McArthur, J.C., 2002. HIV-associated sensory neuropathies. *AIDS,* 16(16), pp. 2105-2117.

Kis, O., Robillard, K., Chan, G.N. and Bendayan, R., 2010. The complexities of antiretroviral drug–drug interactions: role of ABC and SLC transporters. *Trends in Pharmacological Sciences,* **31**(1), pp. 22-35.

Knecht, W., Petersen, G.E., Munch-Petersen, B. and Piškur, J., 2002. Deoxyribonucleoside kinases belonging to the thymidine kinase 2 (TK2)-like group vary significantly in substrate specificity, kinetics and feed-back regulation1. *Journal of Molecular Biology,* **315**(4), pp. 529-540.

Konig, J., Muller, F. and Fromm, M.F., 2013. Transporters and drug-drug interactions: important determinants of drug disposition and effects. *Pharmacological Reviews,* **65**(3), pp. 944-966.

Lane, A.N. and Fan, T.W., 2015. Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Research,* **43**(4), pp. 2466-2485.

Lawn, S.D., Myer, L., Bekker, L.G. and Wood, R., 2006. CD4 cell count recovery among HIV-infected patients with very advanced immunodeficiency commencing antiretroviral treatment in sub-Saharan Africa. *BMC Infectious Diseases*, 6(1), p.59.

Lewis, W., Day, B.J. and Copeland, W.C., 2003. Mitochondrial toxicity of NRTI antiviral drugs: an integrated cellular perspective. *Nature Reviews Drug Discovery,* 2(10), pp. 812-822.

Li, Y., Kang, S., Qin, J., Wang, N., Zhou, R. and Sun, H., 2012. nm23 gene polymorphisms are associated with survival of patients with epithelial ovarian cancer but not with susceptibility to disease. *Gynecologic Oncology,* **126**(3), pp. 455-459.

Life technologies 2015, SNP Genotyping assay, accessed May 7, 2015<https://tools.lifetechnologies.com/content/sfs/brochures/cms_085696.pdf>.

Lockhart, A.C., Tirona, R.G. and Kim, R.B., 2003. Pharmacogenetics of ATP-binding cassette transporters in cancer and chemotherapy. *Molecular Cancer Therapeutics,* **2**(7), pp. 685-698.

Löffler, M., Fairbanks, L.D., Zameitat, E., Marinaki, A.M. and Simmonds, H.A., 2005. Pyrimidine pathways in health and disease. *Trends in Molecular Medicine,* **11**(9), pp. 430-437.

Madian, A.G., Wheeler, H.E., Jones, R.B. and Dolan, M.E., 2012. Relating human genetic variation to variation in drug responses. *Trends in Genetics,* **28**(10), pp. 487-495.

Maekawa, K., Itoda, M., Sai, K., Saito, Y., Kaniwa, N., Shirao, K., Hamaguchi, T., Kunitoh, H., Yamamoto, N., Tamura, T. and Minami, H., 2006. Genetic variation and

haplotype structure of the ABC transporter gene ABCG2 in a Japanese population. *Drug Metabolism and Pharmacokinetics*, 21(2), pp.109-121.

Main, P.A., Angley, M.T., Thomas, P., O'Doherty, C.E. and FenecH, M., 2010. Folate and methionine metabolism in autism: a systematic review. *The American Journal of Clinical Nutrition,* **91**(6), pp. 1598-1620.

Mangravite, L.M., Badagnani, I. and Giacomini, K.M., 2003. Nucleoside transporters in the disposition and targeting of nucleoside analogs in the kidney. *European Journal of Pharmacology,* **479**(1), pp. 269-281.

Maskew, M., Westreich, D., Fox, M.P., Maotoe, T. and Sanne, I.M., 2012. Effectiveness and safety of 30 mg versus 40 mg stavudine regimens: a cohort study among HIV-infected adults initiating HAART in South Africa. *Journal of the International AIDS Society,* **15**(1), pp. 13-2652-15-13.

Massanella, M., Esteve, A., Buzón, M.J., Llibre, J.M., Puertas, M.C., Gatell, J.M., Domingo, P., Stevenson, M., Clotet, B., Martinez-Picado, J. and Blanco, J., 2013. Dynamics of CD8 T-cell activation after discontinuation of HIV treatment intensification. *Journal of Acquired Immune Deficiency Syndromes*, **63**(2), pp.152-160.

May, A., Hazelhurst, S., Li, Y., Norris, S.A., Govind, N., Tikly, M., Hon, C., Johnson, K.J., Hartmann, N., Staedtler, F. and Ramsay, M., 2013. Genetic diversity in black South Africans from Soweto. *BMC genomics*, **14**(1), pp.644.

McArthur, J.C., Brew, B.J. and Nath, A., 2005. Neurological complications of HIV infection. *The Lancet Neurology,* **4**(9), pp. 543-555.

Medicalbiochemistrypage.org 2015, accessed May 7, Synthesis of carbamoyl phosphate, 2015<http://themedicalbiochemistrypage.org/nucleotide-metabolism.php>.

Menezes, C., Duarte, R., Dickens, C., Dix-peek, T., Van Amsterdam, D., John, M., Ive, P., Maskew, M., Macphail, P. and Fox, M., 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), pp. 217-225.

Meyer, U.A., 2000. Pharmacogenetics and adverse drug reactions. *The Lancet,* **356**(9242), pp. 1667-1671.

Michaud, V., Bar-magen, T., Turgeon, J., Flockhart, D., Desta, Z. and Wainberg, M.A., 2012. The dual role of pharmacogenetics in HIV treatment: mutations and polymorphisms regulating antiretroviral drug resistance and disposition. *Pharmacological Reviews,* **64**(3), pp. 803-833.

Miller, D.N., Bryant, J.E., Madsen, E.L. and Ghiorse, W.C., 1999. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Applied and Environmental Microbiology,* **65**(11), pp. 4715-4724.

Minuesa, G., Huber-ruano, I., Pastor-anglada, M., Koepsell, H., Clotet, B. and Martinez-picado, J., 2011. Drug uptake transporters in antiretroviral therapy. *Pharmacology & Therapeutics,* **132**(3), pp. 268-279.

Molina-arcas, M., Casado, F.J. and Pastor-Anglada, M., 2009. Nucleoside transporter proteins. *Current Vascular Pharmacology,* **7**(4), pp. 426-434.

Montier, L.L.C., Deng, J.J. and Bai, Y., 2009. Number matters: control of mammalian mitochondrial DNA copy number. *Journal of Genetics and Genomics,* **36**(3), pp. 125-131.

Moraes, C.T., 2001. What regulates mitochondrial DNA copy number in animal cells? *TRENDS in Genetics,* **17**(4), pp. 199-205.

Müller, P., Asher, N., Heled, M., Cohen, S.B., Risch, A. and Rund, D., 2008. Polymorphisms in transporter and phase II metabolism genes as potential modifiers of the predisposition to and treatment outcome of de novo acute myeloid leukemia in Israeli ethnic groups. *Leukemia Research,* **32**(6), pp. 919-929.

Naicker, N., Kharsany, A.B., Werner, L., van Loggerenberg, F., Mlisana, K., Garrett, N. and Karim, S.S.A., 2015. Risk factors for HIV acquisition in high risk women in a generalised epidemic setting. *AIDS and Behavior*, **19**(7), pp.1305-1316.

Naroueinezhad, M., Khosravi, F., Danesh, A. and Nikbin, B., 2003. Polymerase chain reaction-sequence specific primer (PCR-SSP) versus serology in typing of HLA-a, b and-c in Iranian patients. *ARCHIVES OF IRANIAN MEDICINE* **6**(1), pp. 23-28.

Nasi, M., Borghi, V., Pinti, M., Bellodi, C., Lugli, E., Maffei, S., Troiano, L., Richeldi, L., Mussini, C., Esposito, R. and Cossarizza, A., 2003. MDR1 C3435T genetic polymorphism does not influence the response to antiretroviral therapy in drug-naive HIV-positive patients. *AIDS*, **17**(11), pp.1696-1698.

National antiretroviral treatment guidelines accessed February 23, 2017 http://www.hst.org.za/uploads/files/sa_ART_Guidelines1.pdf.

National Center for Biotechnology Information 2015, deoxythymidylate kinase, accessed May 7, 2015<http://www.ncbi.nlm.nih.gov/gene/1841>.

National Center for Biotechnology Information 2015, deoxyuridine triphosphatase, accessed May 7, 2015< http://www.ncbi.nlm.nih.gov/gene/1854>

National Center for Biotechnology Information 2015, dihydrofolate reductase-like 1, accessed May 7, 2015< http://www.ncbi.nlm.nih.gov/gene/200895>.

National Center for Biotechnology Information 2015, SAM domain and HD domain 1, accessed May 7, 2015<http://www.ncbi.nlm.nih.gov/gene/25939>.

National Center for Biotechnology Information 2015, thymidine kinase 1 gene, accessed May 7, 2015<http://www.ncbi.nlm.nih.gov/gene/7083>.

National guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults accessed February 23, 2017 http://www.sahivsoc.org/Files/ART%20Guidelines%2015052015.pdf.

Nazki, F.H., Sameer, A.S. and Ganaie, B.A., 2014. Folate: metabolism, genes, polymorphisms and the associated diseases. *Gene,* **533**(1), pp. 11-20.

Noguera, N.I., Tallano, C.E., Bragós, I.M. and Milani, A.C., 2000. Modified salting-out method for DNA isolation from newborn cord blood nucleated cells. *Journal of Clinical Laboratory Analysis,* **14**(6), pp. 280-283.

O'Donovan, G.A. and Neuhard, J., 1970. Pyrimidine metabolism in microorganisms. *Bacteriological Reviews,* **34**(3), pp. 278-343.

Oeth, P., Beaulieu, M., Park, C., Kosman, D., Del mistro, G., Van den boom, D. and Jurinke, C., 2005. iPLEX™ assay: increased plexing efficiency and flexibility for MassARRAY® system through single base primer extension with mass-modified terminators. *Sequenom Application Note,* pp. 8876-8006.

Online Mendelian Inheritance in Man 2003, Concentrative nucleoside transporter 1, accessed May 7, 2015<http://omim.org/entry/606207?search=slc28a1&highlight=slc28a1>.

Online Mendelian Inheritance in Man 2007, Concentrative nucleoside transporter 3, accessed May 7, 2015< http://omim.org/entry/608269?search=SLC28A3&highlight=slc28a3>.

Online Mendelian Inheritance in Man 2010, Thymidylate synthetase; TYMS, accessed May 7, 2015<http://omim.org/entry/188350?search=TS%20gene&highlight=gene%20ts>.

Online Mendelian Inheritance in Man 2014, ATP-Binding Cassette, Subfamily C, Member 5; ABCC5, accessed May 7, 2015< http://omim.org/entry/605251?search=ABCC5&highlight=abcc5>.

Online Mendelian Inheritance in Man 2014, NME1-NME2 Spliced read-through transcript, accessed May 7, 2015<http://omim.org/entry/156490?search=NME1-2&highlight=nme12%20nme1>.

Online Mendelian Inheritance in Man 2015, 5, 10-Methylenetetrahydrofolate reductase, accessed May 7, 2015< http://omim.org/entry/607093?search=MTHR&highlight=mthr>.

Oskoui, M., Davidzon, G., Pascual, J., Erazo, R., Gurgel-Giannetti, J., Krishna, S., Bonilla, E., Darryl, C., Shanske, S. and DiMauro, S., 2006. Clinical spectrum of mitochondrial DNA depletion due to mutations in the thymidine kinase 2 gene. *Archives of Neurology*, **63**(8), pp.1122-1126.

Ostermann, N., Schlichting, I., Brundiers, R., Konrad, M., Reinstein, J., Veit, T., Goody, R.S. and Lavie, A., 2000. Insights into the phosphoryltransfer mechanism of human thymidylate kinase gained from crystal structures of enzyme complexes along the reaction coordinate. *Structure,* **8**(6), pp. 629-642.

Pan-Zhou, X.R., Cui, L., Zhou, X.J., Sommadossi, J.P. and Darley-Usmar, V.M., 2000. Differential effects of antiretroviral nucleoside analogs on mitochondrial function in HepG2 cells. *Antimicrobial Agents and Chemotherapy,* **44**(3), pp. 496-503.

Parathyras, J., Gebhardt, S., Hillermann-Rebello, R., Grobbelaar, N., Venter, M. and Warnich, L., 2009. A pharmacogenetic study of CD4 recovery in response to HIV antiretroviral therapy in two South African population groups. Journal of *Human Genetics*, 54(5), pp.261-265.

Pardo, C.A., McArthur, J.C. and Griffin, J.W., 2001. HIV neuropathy: insights in the pathology of HIV peripheral nerve disease. *Journal of the Peripheral Nervous System,* **6**(1), pp. 21-27.

Pastor-Anglada, M. and Pérez-Torras, S., 2015. Nucleoside transporter proteins as biomarkers of drug responsiveness and drug targets. *Frontiers in Pharmacology,* **6**, pp. 13.

Pastor-Anglada, M., Cano-Soldado, P., Molina-Arcas, M., Lostao, M.P., Larráyoz, I., Martínez-Picado, J., et al., 2005. Cell entry and export of nucleoside analogues. *Virus Research,* **107**(2), pp. 151-164.

Pavlos, R. and Phillips, E.J., 2012. Individualization of antiretroviral therapy. *Pharmacogenomics and Personalized Medicine,* **5**, pp. 1.

Pejznochova, M., Tesarova, M., Honzik, T., Hansikova, H., Magner, M. and Zeman, J., 2008. The developmental changes in mitochondrial DNA content per cell in human cord blood leukocytes during gestation. *Physiological Research,* **57**(6), pp. 947.

Phillips, T.J., Brown, M., Ramirez, J.D., Perkins, J., Woldeamanuel, Y.W., Williams, et al., 2014. Sensory, psychological, and metabolic dysfunction in HIV-associated peripheral neuropathy: a cross-sectional deep profiling study. *PAIN®,* **155**(9), pp. 1846-1860.

PIrmohamed, M. and Back, D., 2001. The pharmacogenomics of HIV therapy. *Pharmacogenomics J,* **1**(4), pp. 243-253.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81 PLINK (1.07), Shaun Purcell, http://pngu.mgh.harvard.edu/purcell/PLINK/.

Queiroz, M.A.F., Laurentino, R.V., Amoras, E.D.S.G., de Araújo, M.S.M., Gomes, S.T.M., Lima, S.S., Vallinoto, A.C.R., Ishak, M.D.O.G., Ishak, R. and Machado, L.F.A., 2017. The CYP2B6 G516T polymorphism influences CD4+ T-cell counts in HIV-positive patients receiving antiretroviral therapy in an ethnically diverse region of the Amazon. *International Journal of Infectious Diseases*, *55*, pp.4-10.

Rampazzo, C., Miazzi, C., Franzolin, E., Pontarin, G., Ferraro, P., Frangini, M., et al., 2010. Regulation by degradation, a cellular defense against deoxyribonucleotide pool imbalances. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* **703**(1), pp. 2-10.

Reid, G., Wielinga, P., Zelcer, N., De Haas, M., Van Deemter, L., Wijnholds, J., Balzarini, J. and Borst, P., 2003. Characterization of the transport of nucleoside analog drugs by the human multidrug resistance proteins MRP4 and MRP5. *Molecular Pharmacology,* **63**(5), pp. 1094-1103.

Riddler, S.A., Anderson, R.E. and Mellors, J.W., 1995. Antiretroviral activity of stavudine (2′, 3′-didehydro-3′-deoxythymidine, D4T). *Antiviral Research,* **27**(3), pp. 189-203.

Ross, Philip, Laura Hall, Igor Smirnov, and Larry Haff. "High level multiplex genotyping by MALDI-TOF mass spectrometry." *Nature Biotechnology* 16, no. 13 (1998): 1347-1351.

Rotger, M., Csajka, C. and Telenti, A., 2006. Genetic, ethnic, and gender differences in the pharmacokinetics of antiretroviral agents. Current HIV/AIDS Reports, 3(3), pp.118-125.

Rudin, C.M., Liu, W., Desai, A., Karrison, T., Jiang, X., Janisch, L., Das, S., Ramirez, J., Poonkuzhali, B., Schuetz, E. and Fackenthal, D.L., 2016. Pharmacogenomic and pharmacokinetic determinants of erlotinib toxicity. *Journal of Clinical Oncology*, 26 (2008), pp. 1119-1127.

Rudofsky, G., Schroedter, A., Voron'ko, O.E., Schlimme, M., Tafel, J., Isermann, B.H., et al., 2006. Functional polymorphisms of UCP2 and UCP3 are associated with a reduced prevalence of diabetic neuropathy in patients with type 1 diabetes. *Diabetes Care,* **29**(1), pp. 89-94.

Saada, A., Shaag, A., Mandel, H., Nevo, Y., Eriksson, S. and Elpeleg, O., 2001. Mutant mitochondrial thymidine kinase in mitochondrial DNA depletion myopathy. *Nature Genetics*, **29**(3), p.342.

Saitoh, Akihiko, Elizabeth Sarles, Edmund Capparelli, Francesca Aweeka, Andrea Kovacs, Sandra K. Burchett, Andrew Wiznia, Sharon Nachman, Terence Fenton, and Stephen A. Spector., 2007."CYP2B6 genetic variants are associated with nevirapine pharmacokinetics and clinical response in HIV-1-infected children." *AIDS,* **21**(16), pp. 2191-2199.

Sandrini, M.P. and Piškur, J., 2005. Deoxyribonucleoside kinases: two enzyme families catalyze the same reaction. *Trends in Biochemical Sciences,* **30**(5), pp. 225-228.

Sarfo, F.S., Zhang, Y., Egan, D., Tetteh, L.A., Phillips, R., Bedu-Addo, G., Sarfo, M.A., Khoo, S., Owen, A. and Chadwick, D.R., 2013. Pharmacogenetic associations with plasma efavirenz concentrations and clinical correlates in a retrospective cohort of Ghanaian HIV-infected patients. *Journal of Antimicrobial Chemotherapy*, **69**(2), pp.491-499.

Scholtz, C.L., Odendaal, H.J., Thiart, R., Loubser, L., Hillermann, R., Delport, R., Vermaak, W.H. and Kotze, M.J., 2002. Analysis of two mutations in the MTHFR gene associated with mild hyperhomocysteinaemia–heterogeneous distribution in the South African population. *South African Medical Journal*, **92**(6), pp.464-467.

Schuetz, J.D., Connelly, M.C., Sun, D., Paibir, S.G., Flynn, P.M., Srinivas, R., Kumar, A. and Fridland, A., 1999. MRP4: a previously unidentified factor in resistance to nucleoside-based antiviral drugs. *Nature Medicine,* **5**(9), pp. 1048-1051.

Schuster, S.C., Miller, W., Ratan, A., Tomsho, L.P., Giardine, B., Kasson, L.R., et al., 2010. Complete Khoisan and Bantu genomes from southern Africa. *Nature,* **463**(7283), pp. 943-947.

Schwab, M., Eichelbaum, M. and Fromm, M.F., 2003. Genetic Polymorphisms of the Human MDR1 Drug Transporter. *Annual Review of Pharmacology and Toxicology,* **43**(1), pp. 285-307.

Selvaraj, S., Ghebremichael, M., LI, M., Foli, Y., Langs-barlow, A., Ogbuagu, et al., 2014. Antiretroviral Therapy–Induced Mitochondrial Toxicity: Potential Mechanisms Beyond Polymerase-γ Inhibition. *Clinical Pharmacology & Therapeutics,* **96**(1), pp. 110-120.

Shyamala, V. and Ferro-Luzzi Ames, G., 1993. Single specific primer-polymerase chain reaction (SSP-PCR) and genome walking. PCR Protocols: Current Methods and Applications, pp.339-348.

Sissung, T.M., Baum, C.E., Kirkland, C.T., Gao, R., Gardner, E.R. and Figg, W.D., 2010. Pharmacogenetics of membrane transporters: an update on current approaches. *Molecular Biotechnology,* **44**(2), pp. 152-167.

Staud, F. and Pavek, P., 2005. Breast cancer resistance protein (BCRP/ABCG2). *The International Journal of Biochemistry & Cell Biology*, **37**(4), pp.720-725.

Stein, D.S. and Moore, K.H., 2001. Phosphorylation of nucleoside analog antiretrovirals: a review for clinicians. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy,* **21**(1), pp. 11-34.

Storm, N. and Darnhofer-Patel, B., 2003. MALDI-TOF mass spectrometry-based SNP genotyping. Single nucleotide polymorphisms: *methods and protocols*, **212**(2003), pp.241-262.

Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M.W. and Kubista, M., 2015. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular detection and quantification*, 3, pp.9-16.

Sy, S.K., Innes, S., Derendorf, H., Cotton, M.F. and RosenkranZ, B., 2014. Estimation of intracellular concentration of stavudine triphosphate in HIV-infected children given a reduced dose of 0.5 milligrams per kilogram twice daily. *Antimicrobial Agents and Chemotherapy,* **58**(2), pp. 1084-1091.

Szöke, R.E., Schirmer, M., Gaedcke, J., Brockmöller, J., Becker, H. and Ghadimi, B.M., 2009. Gemcitabine therapy in pancreas carcinoma: Dependence of survival on gene polymorphisms in the nucleoside transporters ENT1 and CNT1. *In Chirurgisches Forum und DGAV,* Forum 2009, pp. 57-59.

Takano, J., Wada, M., Ludewig, U., Schaaf, G., Von Wiren, N. and Fujiwara, T., 2006. The Arabidopsis major intrinsic protein NIP5; 1 is essential for efficient boron uptake and plant development under boron limitation. *The Plant Cell,* **18**(6), pp. 1498-1509.

Tozzi, V., 2010. Pharmacogenetics of antiretrovirals. *Antiviral Research,* **85**(1), pp. 190-200.

Tufi, R., Gandhi, S., De Castro, I.P., Lehmann, S., Angelova, P.R., Dinsdale, D., et al., 2014. Enhancing nucleotide metabolism protects against mitochondrial dysfunction and neurodegeneration in a PINK1 model of Parkinson's disease. *Nature Cell Biology,* **16**(2), pp. 157-166.

Uchida, K., Hayashi, K., Kawakami, K., Schneider, S., Yochim, J.M., Kuramochi, H., Takasaki, K., Danenberg, K.D. and Danenberg, P.V., 2004. Loss of heterozygosity at the thymidylate synthase (TS) locus on chromosome 18 affects tumor response and survival in individuals heterozygous for a 28-bp polymorphism in the TS gene. *Clinical Cancer Research: an official Journal of the American Association for Cancer Research,* **10**(2), pp. 433-439.

Uhlin, U. and Eklund, H., 1994. Structure of ribonucleotide reductase protein R1. *Nature*, *370*(6490), pp.533-539.

Urata, M., Koga-Wada, Y., Kayamori, Y. and Kang, D., 2008. Platelet contamination causes large variation as well as overestimation of mitochondrial DNA content of peripheral blood mononuclear cells. *Annals of Clinical Biochemistry*, **45**(5), pp.513-514.

Van Oosterhout, J.J., Mallewa, J., Kaunda, S., Chagoma, N., Njalale, Y., Kampira, et al., 2012. Stavudine toxicity in adult longer-term ART patients in Blantyre, Malawi. *PLoS One,* **7**(7), pp. e42029.

Van Rompay, A.R., Johansson, M. and Karlsson, A., 2000. Phosphorylation of nucleosides and nucleoside analogs by mammalian nucleoside monophosphate kinases. *Pharmacology & Therapeutics,* **87**(2), pp. 189-198.

Vasiliou, V., Vasiliou, K. and Nebert, D.W., 2009. Human ATP-binding cassette (ABC) transporter family. *Human Genomics,* **3**(3), pp. 281-290.

Verma, S. and Simpson, D.M., Peripheral neuropathy in HIV infection. *Handbook of Clinical Neurology.* Elsevier, pp. 129-137.

Visscher H, Ross CJ, Rassekh SR, Sandor GS, Caron HN, van Dalen EC, Kremer LC, van der Pal HJ, Rogers PC, Rieder MJ, Carleton BC., 2013. Validation of variants in SLC28A3 and UGT1A6 as genetic markers predictive of anthracycline-induced cardiotoxicity in children. *Pediatric Blood & Cancer*, **60**(8), pp. 1375-81.

Visscher, H., Ross, C.J., Rassekh, S.R., Barhdadi, A., Dubé, M.P., Al-Saloos, H., Sandor, G.S., Caron, H.N., van Dalen, E.C., Kremer, L.C. and van der Pal, H.J., 2011.

Pharmacogenomic prediction of anthracycline-induced cardiotoxicity in children. *Journal of Clinical Oncology*, **30**(13), pp.1422-1428.

Wadley, A., Kamerman, P., Chew, C., Lombard, Z., Cherry, C. and Price, P., 2013. A polymorphism in IL4 may associate with sensory neuropathy in African HIV patients. *Molecular Immunology,* **55**(3), pp. 197-199.

Wadley, A.L., Cherry, C.L., Price, P. and Kamerman, P.R., 2011. HIV neuropathy risk factors and symptom characterization in stavudine-exposed South Africans. *Journal of Pain and Symptom Management,* **41**(4), pp. 700-706.

Wadley, A.L., Hendry, L.M., Kamerman, P.R., Chew, C.S., Price, P., Cherry, C.L. and Lombard, Z., 2014. Role of TNF block genetic variants in HIV-associated sensory neuropathy in black Southern Africans. *European Journal of Human Genetics*, doi: 10.1038/ejhg.2014.104.

Wadley, A.L., Lombard, Z., Cherry, C.L., Price, P. and Kamerman, P.R., 2013. Polymorphisms in uncoupling protein genes UCP2 and UCP3 are not associated with HIV-associated sensory neuropathy in African individuals. *Journal of the Peripheral Nervous System,* **18**(1), pp. 94-96.

Wangsomboonsiri, W., Mahasirimongkol, S., Chantarangsu, S., Kiertiburanakul, S., Charoenyingwattana, A., Komindr, S., et al., 2010. Association between HLA-B*4001 and lipodystrophy among HIV-infected patients from Thailand who received a stavudine-containing antiretroviral regimen. *Clinical Infectious Diseases: an official Publication of the Infectious Diseases Society of America,* **50**(4), pp. 597-604.

Wegner, S., 2002. Racial differences in clinical efficacy of efavirenz-based antiretroviral therapy. In Ninth Conference on Retroviruses and Opportunistic Infections (Seattle, WA February 428-W, 2002).

Winzer, R., Langmann, P., Zilly, M., Tollmann, F., Schubert, J., Klinker, H. and Weissbrich, B., 2005. No influence of the P-glycoprotein polymorphisms MDR1 G2677T/A and C3435T on the virological and immunological response in treatment naive HIV-positive patients. *Annals of Clinical Microbiology and Antimicrobials*, **4**(1), p.3.

Wojtal, K.A., Eloranta, J.J., Hruz, P., Gutmann, H., Drewe, J., Staumann, A., et al., 2009. Changes in mRNA expression levels of solute carrier transporters in inflammatory bowel disease patients. *Drug Metabolism and Disposition: the Biological Fate of Chemicals,* **37**(9), pp. 1871-1877.

Wong, A., Soo, R.A., Yong, W.P. and Innocenti, F., 2009. Clinical pharmacology and pharmacogenetics of gemcitabine. *Drug Metabolism Reviews*, **41**(2), pp.77-88.

Yang, G., Dutschman, G.E., Wang, C., Tanaka, H., Baba, M., Anderson, K.S. and Cheng, Y., 2007. Highly selective action of triphosphate metabolite of 4′-ethynyl D4T: a novel anti-HIV compound against HIV-1 RT. *Antiviral Research,* **73**(3), pp. 185-191.

Yu, K., Yoon, T., Minai-Tehrani, A., Kim, J., Park, S.J., Jeong, M.S., Ha, S., Lee, J., Kim, J.S. and Cho, M., 2013. Zinc oxide nanoparticle induced autophagic cell death and mitochondrial damage via reactive oxygen species generation. *Toxicology in Vitro,* **27**(4), pp. 1187-1195.

Yuan, J.S., Reed, A., Chen, F. and Stewart, C.N., 2006. Statistical analysis of real-time PCR data. *BMC bioinformatics*, *7*(1), p.85.

Zhang, Z., Yamashita, H., Toyama, T., Yamamoto, Y., Kawasoe, T. and Iwase, H., 2006. Reduced expression of the breast cancer metastasis suppressor 1 mRNA is correlated with poor progress in breast cancer. *Clinical cancer research*, *12*(21), pp.6410-6414.

APPENDIX A: ETHICAL CLEARANCE

R14/49 Dr Debra de Assis

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M150459

NAME: (Principal Investigator)	Dr Debra de Assis
DEPARTMENT:	School of Molecular and Cell Biology
PROJECT TITLE:	Associations between Variation in Candidate Genes and Sensory Neuropathy in South Africans using d4T
DATE CONSIDERED:	24/04/2015
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	
APPROVED BY:	Professor P Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	22/05/2015

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Secretary in Room 10004, 10th floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**.

Principal Investigator Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

APPENDIX B: GENOTYPE FREQUENCIES OF SA POPULATION.

Chr[a]	SNP ID	Genotypes	Genotypic SN-free	Genotypic SN	Genotypic frequency (All)	HWE (All) p-value
d4T transporters						
3	rs3749442	AA	0.0183	0.0522	0.0370	0.697
		AG	0.3303	0.3507	0.3416	
		GG	0.6514	0.5970	0.6214	
4	rs2725252	AA	0.0367	0.0299	0.0329	0.837
		AC	0.3211	0.3209	0.321	
		CC	0.6422	0.6492	0.6461	
4	rs2622604	TT	0.0185	0.0229	0.0209	0.575
		TC	0.2037	0.2366	0.2218	
		CC	0.7778	0.7405	0.7573	
4	rs12505410	GG	0.0093	0.0149	0.0124	1.000
		GT	0.1776	0.2537	0.2199	
		TT	0.8131	0.7313	0.7676	
4	rs3114018	CC	0.1354	0.1832	0.1630	0.497
		CA	0.5104	0.5115	0.511	
		AA	0.3542	0.3053	0.3260	
9	rs4877847	CC	0.2037	0.2313	0.2190	1.000
		CA	0.5463	0.4627	0.5	
		AA	0.25	0.3060	0.2810	
9	rs7853758	AA	0.2385	0.0379	0.2428	0.701
		AG	0.5046	0.5224	0.5144	
		GG	0.2569	0.7652	0.2428	
15	rs2242046	AA	0.000	0.000	0.000	0.575
		AG	0.0280	0.06716	0.0498	
		GG	0.9720	0.9328	0.9502	
15	rs2290272	AA	0.046	0.05344	0.05	0.597
		AG	0.4495	0.3282	0.3833	
		GG	0.5046	0.6183	0.5667	

15	rs8187758	AA	0.0093	0.0379	0.0251	0.808
		AC	0.3364	0.197	0.2594	
		CC	0.6542	0.7652	0.7155	
dNTP regulators						
1	rs1801131	GG	0.02752	0.0075	0.0165	0.795
		GT	0.2752	0.2388	0.2551	
		TT	0.6972	0.7537	0.7284	
1	rs1801133	AA	0.0	0.0075	0.0041	1.000
		AG	0.1376	0.1269	0.1317	
		GG	0.8624	0.8657	0.8642	
2	rs7574663	GG	0.0092	0.0149	0.0082	0.680
		GC	0.1651	0.1418	0.1523	
		CC	0.8257	0.8507	0.8395	
5	rs1650723	TT	0.0	0.0	0.000	0.608
		TC	0.1284	0.1194	0.1235	
		CC	0.8716	0.8806	0.8765	
8	rs16918482	CC	0.0092	0.0075	0.0083	0.249
		CA	0.1193	0.1053	0.1116	
		AA	0.8716	0.8806	0.8802	
11	rs1465952	AA	0.2569	0.2331	0.2438	0.306
		AG	0.4495	0.4812	0.4669	
		GG	0.2936	0.2857	0.2893	
11	rs11030918	CC	0.0183	0.0526	0.037	1.000
		CT	0.3303	0.3233	0.3264	
		TT	0.6514	0.6241	0.6364	
11	rs12806698	AA	0.000	0.000	0.000	1.000
		AC	0.03704	0.0448	0.04132	
		CC	0.9630	0.9552	0.9587	
11	rs1042927	CC	0.0092	0.0149	0.0123	0.032
		CA	0.3578	0.3209	0.3374	
		AA	0.6330	0.6642	0.6502	
16	rs3743712	GG	0.1667	0.1716	0.1694	0.281

		GA	0.4815	0.4104	0.4421	
		AA	0.3519	0.4179	0.3884	
16	rs11859474	AA	0.0642	0.0970	0.0823	0.321
		AG	0.4037	0.3284	0.3621	
		GG	0.4771	0.5746	0.5556	
16	rs2288399	GG	0.0459	0.0299	0.0370	0.130
		GA	0.2202	0.2388	0.2305	
		AA	0.7339	0.7313	0.7325	
20	rs8124728	AA	0.0556	0.0149	0.0330	0.647
		AG	0.213	0.3209	0.2727	
		GG	0.7315	0.6642	0.6942	
20	rs1291142	GG	0.0459	0.0602	0.0537	1.000
		GA	0.4128	0.3083	0.3554	
		AA	0.5413	0.6316	0.5885	
20	rs1891643	GG	0.0183	0.0448	0.0329	1.000
		GA	0.3394	0.2836	0.3086	
		AA	06422	0.6716	0.6584	
21	rs1051266	CC	0.0734	0.0896	0.0823	0.369
		CT	0.4128	0.4925	0.4568	
		TT	0.5138	0.4179	0.4609	

APPENDIX C: MASS EXTEND EXTENSION PRIMERS.

Locus	PCR primer sequence	Length	Tm (°C)
rs2622604_W1_F	ACGTTGGATGACTCTGAAAGCACTGTTTTG	30	66.15
rs2622604_W1_R	ACGTTGGATGCATTTGAATGTCAGCTAGTC	30	65.92
rs2725252_W1_F	ACGTTGGATGTTAGCCAAGCAGTCAGGAAC	30	68.72
rs2725252_W1_R	ACGTTGGATGGCTGATTGAAGGTGCATAAC	30	67.58
rs4877847_W2_F	ACGTTGGATGGATCCCTGATGGACTTACAC	30	67.68
rs4877847_W2_R	ACGTTGGATGCCATGAGTAAGCCAGTCAAG	30	68.60
rs2242046_W1_F	ACGTTGGATGAGACCTCTCCAAGTACAAGC	30	67.76
rs2242046_W1_R	ACGTTGGATGACACTCACGGAGATCCACTG	31	69.46
rs3114018_W1_F	ACGTTGGATGGCCACTGAGGAAAAGACTTC	30	68.53
rs3114018_W1_R	ACGTTGGATGGGAAACCTCACAAAAGTGTC	30	67.41
rs2290272_W2_F	ACGTTGGATGTGAGCAACTGGTGTCCTTCG	30	70.27
rs2290272_W2_R	ACGTTGGATGCTGCGCAATGATGCTTTGAG	30	70.12
rs3749442_W2_F	ACGTTGGATGGGTTTGATCGTGAAGGATAC	30	66.78
rs3749442_W2_R	ACGTTGGATGTGACCTTTGAGAACGCAGAG	30	68.65
rs1650723_W1_F	ACGTTGGATGCCATGGAGATATATTGCCATC	31	66.83
rs1650723_W1_R	ACGTTGGATGCACTTGTTAATAAGTGCTGG	30	66.08
rs1042927_W1_F	ACGTTGGATGCAGGGAGTGGTTAAGTAAGG	30	68.00
rs1042927_W1_R	ACGTTGGATGTAGAGACCCACCAGTCAAAG	30	67.74
rs8187758_W2_F	ACGTTGGATGTACTTGGACTCCTCGTCATC	30	67.58

rs8187758_W2_R	ACGTTGGATGCCAGACCCCACATACCTAC	30	69.42
rs7853758_W1_ F	ACGTTGGATGTCAGCTGTGGGTAGTCAAAC	30	68.48
rs7853758_W1_ R	ACGTTGGATGCCAACATCGCTGTGAATCTG	30	69.34
rs1801131_W2_F	ACGTTGGATGTCTCCCGAGAGGTAAAGAAC	30	67.56
rs1801131_W2_R	ACGTTGGATGTCTACCTGAAGAGCAAGTCC	30	67.76
rs1801133_W2_F	ACGTTGGATGGTGCATGCCTTCACAAAGCG	30	71.49
rs1801133_W2_R	ACGTTGGATGCACTTGAAGGAGAAGGTGTC	30	68.50
rs12806698_W1_F	ACGTTGGATGACTCAACATGGCGGCTACAC	30	70.35
rs12806698_W1_R	ACGTTGGATGAATGACGTTACTCGACGCTG	30	68.65
rs12505410_W2_F	ACGTTGGATGCCTTGGCACCTTAAATGAAC	30	67.93
rs12505410_W2_R	ACGTTGGATGCATGATTCAAACTTGGCCTC	30	67.96
rs8124728_W1_F	ACGTTGGATGCTCCAACTCTTTACCCCAAC	30	68.55
rs8124728_W1_R	ACGTTGGATGTCTGGCTCTACAGGGATGC	30	69.60
rs1891643_W1_ F	ACGTTGGATGCATTGTCAGTACTCCCTCAC	30	68.17
rs1891643_W1_R	ACGTTGATGGCAGGATAATGAATCACAGG	29	65.37
rs3743712_W2_F	ACGTTGGATGTCTGGCTCTACAGGGATGC	29	68.55
rs3743712_W2_R	ACGTTGGATGAGATGGCACATGGGCACAAC	30	69.60
rs16918482_W2_F	ACGTTGGATGCTTGGCAAACTGAAGCAATG	30	68.65
rs16918482_W2_R	ACGTTGGATGCAAAAGGAGCAGGCCATAAC	30	69.41
rs1465952_W2_F	ACGTTGGATGTCCTTTGCCCAGGGAAAATG	30	69.21
rs1465952_W2_R	ACGTTGGATGTTTATCTGGTAACCCTGCCC	30	68.26
rs11030918_W1_F	ACGTTGGATGACCATCCACTGAAGAACCCT	30	69.06

rs11030918_W1_R	ACGTTGGATGTCCTGACGCAAATCAGAGCC	30	70.37
rs7574663_W1_F	ACGTTGGATGGCAAATGTAGAATGTGGGAC	30	67.37
rs7574663_W1_R	ACGTTGGATGCTTTTCCCCACGACATTGAC	30	69.30
rs11859474_W2_F	ACGTTGGATGTTACTCCAACCAGCGAGTGA	30	69.33
rs11859474_W2_R	ACGTTGGATGTTCGTGTCCTCATGCAGTTC	30	68.87
rs1051266_W1_ F	ACGTTGGATGTGAAGCCGTAGAAGCAAAGG	30	68.91
rs1051266_W1_ R	ACGTTGGATGAGAAGCAGGTGCCCGTGGAA	30	72.87
rs1291142_W1_ F	ACGTTGGATGAGCTCATTCCAGCTTCATGC	30	69.03
rs1291142_W1_ R	ACGTTGGATGGTGTTTGGATGAGTAGCAGG	30	68.39
rs2288399_W1_ F	ACGTTGGATGGAAGGACAGCTCAAGAAACC	30	68.53
rs2288399_W1_ R	ACGTTGGATGTCAGAGCCCTGTGATCATGC	30	70.07