

https://dspace.org/

Genetic variation influencing mitochondrial DNA copy number and the development of sensory neuropathy in HIV-positive patients exposed to stavudine

Marutha, Tebogo Rector

2017

https://hdl.handle.net/10539/24162

Marutha, Tebogo Rector (2017) Genetic variation influencing mitochondrial DNA copy number and the development of sensory neuropathy in HIV-positive patients exposed to stavudine, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/24162>

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

GENETIC VARIATION INFLUENCING MITOCHONDRIAL DNA COPY NUMBER AND THE DEVELOPMENT OF SENSORY NEUROPATHY IN HIV-POSITIVE PATIENTS EXPOSED TO STAVUDINE.

BY

MARUTHA TR

1111612

SUPERVISOR: Dr D. De Assis Rosa

August 2017

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree in Master of Science in the School of Molecular and Cell Biology.

DECLARATION

I, Tebogo Rector Marutha (1111612), am a student registered for the degree of Masters of Science in Genetics (Dissertation) 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_T.R Marutha_ day of_01/08/_2017__ **YEAR**

I would like to dedicate this work to my loving family for all the support they have given me.

ABSTRACT

Antiretroviral therapy (ART) drugs such as stavudine (d4T) are known to have off-target side-effects, including the inhibition of DNA polymerase gamma which replicates mitochondrial DNA (mtDNA). ART-induced depletion of mtDNA copy number may cause mitochondrial toxicities such as sensory neuropathy (SN). Genetic variation in DNA polymerase gamma or in other nuclear genes influencing mtDNA replication and mtDNA copy number may therefore contribute to susceptibility to d4T-induced SN. DNA samples from 263 HIV-positive South African adults exposed to d4T were classified as cases with SN (n = 143) and controls without SN (n = 120). A total of 28 single nucleotide polymorphism (SNPs) were chosen in nuclear genes from the mtDNA replication pathway and from a GWAS paper examining SNP association with ART-induced SN (Leger et al. 2014). Genotyping was performed using Sequenom Mass Spectrometry. MtDNA copy number was determined using a qPCR assay. Associations between SN and genetic variants, between genetic variants and mtDNA copy number, and between mtDNA copy number and SN were evaluated in univariate and multivariate models using Plink v1.07 and GraphPad v7. Age and height were significantly different in the cases with SN vs controls without SN. In univariate analyses, three SNPs and two haplotypes were significantly associated with SN, three SNPs were associated with pain intensity and three haplotypes were significantly associated with mtDNA copy number. However, there were no significant associations with SN, pain intensity or mtDNA copy number after correction for multiple SNP testing. No significant difference in mtDNA copy number in cases vs. controls was observed. In conclusion variation in nuclear-encoded mitochondrial genes examined in the current study do not play a role in ART-related mitochondrial complications such as changes in mtDNA copy number, or occurrence of SN.

ACKNOWLEDGEMENTS

I would like to give thanks to the listed people for their contributions to the success of this work:

- Dr Debra de Assis Rosa (supervisor) and Prof Monde Ntwasa (advisor) for their perseverance, technical supervision and the knowledge of which they imparted to me.
- Cohort kindly supplied by Dr Antonia Wadley and Prof P Kamerman, School of Physiology, University of the Witwatersrand.
- Voluntary participants in the study
- Tyrone Otgaar and Dr Boitelo Letsolo for assistance with cell culture
- Marvin Moketla for assistance with qPCR
- My parents Mr Headman Marutha and Mrs Agnes Marutha, for their support and encouragement in all aspects of my life.
- Staff members for their willingness to help me at all times.
- The Gate House 700 students for their encouragement and the lovely years that we had together.
- The University of the Witwatersrand for giving me this opportunity
- National Research Foundation (NRF) Innovation Master's scholarship for financial support.

And over and above I thank the All Mighty GOD for everything he has made possible.

CONFERENCE PRESENTATIONS ASSOCIATED WITH THIS STUDY

Marutha, TR., Moketla, MB., 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 [16-19 August 2015].

Marutha, TR., Wadley, A., Kamerman, P., de Assis Rosa, D. 2016. "*NRF2* genetic variation may confer protective effect against sensory neuropathy in HIV-positive South Africans using d4T". Poster Presentation at the Molecular Biosciences Research Thrust (MBRT) Research day at Wits University [08 December 2016].

TABLE OF CONTENTS

DECLARATION i
ABSTRACT iii
ACKNOWLEDGEMENTS iv
CONFERENCE PRESENTATIONS ASSOCIATED WITH THIS STUDY v
TABLE OF CONTENTS vi
LIST OF FIGURES ix
LIST OF TABLES xi
ABBREVIATIONS xiii
CHAPTER 1: INTRODUCTION 1
1.1 HIV epidemic and HAART 2
1.2 Sensory neuropathy 4
1.4 Mitochondrial DNA copy number and ART side-effects 10
1.5 Pharmacogenetics of ART-induced sensory neuropathy 13
1.6 Genetics of mitochondrial DNA copy number 16
AIMS AND OBJECTIVES 24
1.7 Hypotheses and Aims 25
1.8 Objectives 26
CHAPTER 2: MATERIALS AND METHODS 27
2.1 Case-control cohort 28
2.1.1 Sample collection 28
2.1.2 Phenotypic data collection and SN diagnosis 28
2.1.3 DNA isolation 29
2.2 mtDNA copy number assessment 29
2.2.1 PCR Optimisation 30
2.2.2 qPCR assay validation 37
2.2.3 mtDNA copy number in cohort samples 39
2.3 SNP choice (Bioinformatics analysis) 41
2.3.1 Literature review of candidate genes 41
2.3.2 Minor allele frequency (MAF) 42

2.3.3 Predicted functional effect 42
2.3.4 SNP assay in Sequenom 42
2.3.5 Tag SNPs 43
2.4 SNP genotyping 44
2.5 Statistical analysis 48
2.5.1 Analysis of association between demographic or clinical data, and SN 48
2.5.2 Analysis of association between demographic or clinical data, and mtDNA copy number 48
2.5.3 Quality control (QC) of SNP data 48
2.5.4 Analysis of allele, genotype and haplotype frequencies in the whole cohort 49
2.5.5 Analysis of SNP associations with phenotypes 49
2.5.5.1 Univariate analysis 51
2.5.6 Genetic risk score (GRS) 53
CHAPTER 3: RESULTS 55
3.1 Association of demographic characteristics with SN 56
3.2 Cohort DNA quantification 57
3.3 qPCR assay 58
3.3.1 Conventional PCR and primer optimisation 58
3.3.2 qPCR optimisation 60
3.3.3 qPCR assay validation 61
3.3.4 qPCR in cohort samples 63
3.4 SNP selection in genes of interest. 65
3.5 Sequenom genotyping results 70
3.6 Allele frequencies in whole cohort vs. other African populations 72
3.7 Haplotypes and linkage disequilibrium (LD) in cohort 74
3.7.1 Haplotype frequencies in cohort 74
3.7.2 LD analysis 78
3.8 Associations between alleles, genotypes, haplotypes and SN. 79
3.8.1 Univariate analysis of associations between genetic variation and SN 79
3.8.2 Multivariate analysis of association with SN 90
3.9 Associations between alleles, genotype, haplotypes and pain intensity 93
3.9.1 Univariate Associations between genetic variants and pain intensity 93

3.9.2 Multivariate analysis of genetic associations with pain intensity.....103
3.10 Associations between alleles, genotype, haplotypes and mtDNA copy number106
3.10.1 Univariate analysis of associations between genetic variants and mtDNA copy number.....106
3.10.2 Multivariate analysis of genetic variants vs mtDNA copy number.....111
3.11 Genetic risk scores (GRS).....113
3.11.1 GRS and SN.....113
3.11.2 GRS and pain intensity.....114
3.11.3 GRS and mtDNA copy number.....115
CHAPTER 4: DISCUSSION AND CONCLUSION 116
4.1 Genetic variation in nuclear genes of the mtDNA replication pathway in the black SA population.....119
4.2 Association of demographic factors with SN and with mtDNA copy number.....119
4.3 Genetic associations with SN after d4T use.....120
4.4 Genetic associations with pain intensity after d4T use.....123
4.5 mtDNA copy number association with SN 125
4.6 Genetic associations with mtDNA copy number after d4T use 125
REFERENCES 131
APPENDICES.....148
APPENDIX A 148
APPENDIX B 149
APPENDIX C 150

LIST OF FIGURES

Figure 1.1 Average mtDNA: nDNA ratios in peripheral blood samples……. …… **7**
obtained from different groups of HIV infected patients

Figure 1.2 Effect of antiretroviral therapy with or without d4T on median……..... **8**
mtDNA amounts in HIV-negative and HIV-positive patients

Figure 1.3 Depletion of mitochondrial DNA (mtDNA) in skeletal muscle cells ….. **9**
associated with specific NRTIs

Figure 1.4 A model of two ways in which NRTIs may induce neuropathic pain in ….. **10**
HIV patients on ART

Figure 1.5 Mitochondrial DNA replication fork…….. ….. **24**

Figure 1.6 Genotype/phenotype associations and phenotype/phenotype…….. ….. **25**
associations (blue arrows) that will be examined in this study

Figure 2.1 qPCR quantification of DNA using SYBR Green I for detection….. …. **30**

Figure 2.2 High throughput SNP genotyping work-flow with MassARRay……. **47**
Spectrophotometer

Figure 2.3 Model for analysis of SNP data with the aid of Plink…… …. **50**

Figure 3.1 Optimisation results of *GAPDH* primer set 2…… …. **59**

Figure 3.2 Optimisation results of *GAPDH* primer set 1…… …. **59**

Figure 3.3 Optimisation results of *NADH1* primers. (A) Primer pair 2, (B) Primer …. **59**
pair 1

Figure 3.4 qPCR primer optimisation of target and reference genes……..... …. **60**

Figure 3.5 Comparisons of average Ct values of *NADH1* & *GAPDH* genes…..... …. **61**

Figure 3.6 mtDNA copy number in cases vs controls……..... …. **64**

Figure 3.7 …. **70**
High throughput genotyping data with the MassARRay system

showing SNPs distributed across the range of detectable DNA masses from 4500 to 9000 Da

Figure 3.8 Linkage Disequilibrium plot of the interrogated SNP markers…….. …. **78**

Figure 3.9 GRS vs SN…….. … **114**

Figure 3.10 GRS vs pain intensity in patients with SN…….. … **115**

Figure 3.11 GRS vs mtDNA copy number in HIV-positive patients exposed to…...… **116** d4T

LIST OF TABLES

Table 1.1 Adverse effects associated with NRTIs and the reported manifestations of mitochondrial toxicity **…. 3**

Table 1.2 Genes already examined in association with SN in various cohorts **…… 15**

Table 1.3 Proteins involved in mtDNA transcription & help initiate DNA replication **…… …. 19**

Table 1.4 Proteins involved in mtDNA replication **……. … 22**

Table 2.1 nDNA & mtDNA qPCR primers and conditions **…….....… 32**

Table 2.2 Kapa Taq PCR reaction mix **…… … 33**

Table 2.3 Kapa Taq Cycling conditions **……...… 34**

Table 2.4 Annealing tTemperature rRanges **…… … 35**

Table 2.5 Thermal cycling conditions as from the SYBR master mix protocol **.. … 36**

Table 2.6 qPCR reaction mix utilised for quantification of mtDNA **……... … 36**

Table 2.7 Cell culture, treatment and harvest schedule **…… 38**

Table 2.8 Template of the 96-well plate map used in this study **……..…. 40**

Table 2.9 SNPs & their corresponding plex number during genotyping on the MassArray spectrophotometer **… 44**

Table 2.10 Reaction Components for PCR of Target Regions **…….. … 45**

Table 2.11 Thermal Cycler Conditions for Amplification of Target Regions **….... … 45**

Table 2.12 Reaction Components for Single Base Extension Reaction **……... … 46**

Table 2.13 Thermal Cycling conditions of MassARRay instrument using iPlex Gold reaction mix **… 46**

Table 2.14 Risk alleles and non-risk alleles of SNPs of interest **…… 54**

Table 3.1 Demographic characteristics of study participants **…….. … 56**

Table 3.2 Association of demographic data with mtDNA copy number **……. … 57**

Table 3.3 qPCR threshold cycle (Ct) values obtained using HEK cells DNA **…. … 61**

Table 3.4 Fold change ratios of *NADH1* relative to *GAPDH* **……... … 62**

Table 3.5 Average $2^{\wedge dCt}$ fold difference in cases vs controls and exact P-value **...…. 63**

Table 3.6 List of SNPs chosen for genotyping in this cohort **……... . 66-69**

Table 3.7 Plink QC data of whole cohort samples **…….....…. 71**

Table 3.8 Distribution of allele frequencies in three African population groups, two proxy populations [Kenya (LWK) and Nigeria (YRI)] **… 73**

and South Africans (Blacks)

Table 3.9 Frequencies of 3 SNP haplotypes in whole cohort……... **75-77**

Table 3.10 Test of association of alleles with SN ……..... **80-81**

Table 3.11 Association of genotypes with SN using the full association test models **82-85**

Table 3.12 Haplotype association (Univariate) with SN……... **87-90**

Table 3.13 Logistic regression analysis of association between alleles and SN including age and height as covariates **… … 91**

Table 3.14 Logistic regression analysis of association between genotypes and SN including age and height as covariates **.. … 91**

Table 3.15 Additive effect of age and height to haplotype association with SN (Logistic regression analysis) **… 92**

Table 3.16 Summary of results from association of genetic variants with SN……... **… 93**

Table 3.17 Basic association test of alleles with the phenotype pain intensity..... **… 95**

Table 3.18 Association of genotypes with pain intensity using the full association test models **96-99**

Table 3.19 Haplotype association with the phenotype pain intensity……. **101-103**

Table 3.20 Linear regression analysis of alleles in association with pain intensity **…….. ... 104**

Table 3.21 Linear regression analysis of genotypes in association with Pain Intensity **…… .. 105**

Table 3.22 Summary of results from association of genetic variants with pain intensity **..... .. 106**

Table 3.23 Basic association test of alleles with the phenotype mtDNA copy number **….. .. 108**

Table 3.24 Haplotype Association with mtDNA copy number…….. **109-111**

Table 3.25 Summary of results from association of genetic variants with mtDNA copy number **……. 113**

ABBREVIATIONS

3TC	Lamivudine
ABC	Abacavir
AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Therapy
ARV	Antiretroviral Therapy
ATN	Antiretroviral Toxic Neuropathy
ATP	Adenosine Triphosphate
AZT	Zidovudine
C10orf2	Twinkle Helicase
CD4	Cluster of Differentiation 4
CT	Threshold Cycle
d4T	Stavudine
ddc	Zalcitabine
DDI	Didanosine
ΔΔCT	Delta Delta Threshold Cycle
D-loop	Displacement Loop
DM1	Type 1 Diabetes Mellitus
DNA	Deoxyribonucleic Acid
DPN	Diabetic Polyneuropathy
DRG	Dorsal Root Ganglion
DRG	Dorsal Root Ganglion
dsDNA	Double stranded Deoxyribonucleic Acid
DSP	Distal Symmetric Polyneuropathy
DSPN	Distal Symmetric Peripheral neuropathy
EMP	Empirical P-value
EtBr	Ethidium Bromide
FDC	Fixed Dose Combination
FTC	Emtricitabine
GCH1	GTP Cyclohydrolase 1
GRS	Genetic Risk Score
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Virus
IL-1	Interleukin-1
IL-12	Interleukin-12
IL-6	Interleukin-6
LD	Linkage Disequilibrium
LPV	Lopinavir
LPV/r	Lopinavir / Ritonavir

LWK	Luhya in Webuye Kenya
MAF	Minor Allele Frequency
MDS	Mitochondrial Depletion Syndrome
MERRF	Myoclonic Epilepsy With Ragged-Red Fibres
MHC	Major Histocompatibility Complex
MIQE	Minimum Information for Publication of Quantitative Real-Time PCR Experiments
MNGIE	Mitochondrial Neurogastrointestinal Encephalopathy
mtDNA	Mitochondrial DNA
NA	Not Available
nDNA	Nuclear DNA
NG	Not Genotyped
NIEHS	National Institute of Environmental Health Sciences
NIH PDR	National Institute of Health Polymorphism Discovery Resource
NNRTI	Non-Nucleoside Reverse Transcriptase Inhibitor
NG	Nanograms
NM	Nanometre
NRF1	Nuclear Respiratory Factor 1
NRF2	Nuclear Respiratory Factor 2
NRF3	Nuclear Respiratory Factor 3
NRTI	Nucleoside Reverse Transcriptase Inhibitor
NVP	Nevaripine
OR	Odds Ratio
PBMC	Peripheral Blood Mononuclear Cells
PCR	Polymerase Chain Reaction
pEMP1	Empirical P-value
pEMP2	Empirical P-value
PEO	Progressive External Opthalmoplegia
PGC-1a	Peroxisome Proliferator-Activated Receptor Gamma Coactivator 1-Alpha
PI	Protease Inhibitors
PINK1	PTEN-Induced Putative Kinase 1
PN	Peripheral Neuropathy
POLG	DNA Polymerase Gamma
QC	Quality Control
qPCR	Quantitative Polymerase Chain Reaction
ROS	Reactive Oxygen Species
rRNA	Ribosomal Ribonucleic Acid
RS ID	Reference SNP Cluster ID

SANDO	Sensory Ataxia Neuropathy, Dysarthia and Opthalmoplegia
SIFT	Sorting Intolerant From Tolerant
SN	Sensory Neuropathy
SNP	Single Nucleotide Polymorphism
ssDNA	Single Stranded Deoxyribonucleic Acid
TDF	Tenofovir Disoproxil Fumarate
Ta	Annealing Temperature
Tm	Melting Temperature
TNFA	Tumour Necrosis Factor A
tRNA	Transfer Ribonucleic Acid
UCP	Mitochondrial Uncoupling Proteins
µL	Microliter
µM	Micrometre
UNAIDS	The Joint United Nations Programme on HIV and AIDS
VEP	Variant Effect Predictor
Vs	Versus
WHO	World Health Organisation
YRI	Yoruba

CHAPTER 1: INTRODUCTION

1.1 HIV epidemic and HAART

Over the past 30 years, Human Immunodeficiency Virus (HIV) and the resulting Acquired Immunodeficiency Syndrome (AIDS) has caused a globally devastating pandemic (Streatfield *et al.,* 2014; Singh *et al.,* 2012) with an estimated 1.5 million deaths in 2013 (Murray *et al.,* 2014). The HIV/AIDS epidemic remains a prominent health concern in Africa, particularly in South Africa with an estimated population of 5.51 million living with the disease in 2014 and over 170 700 deaths in 2014 (Statistics South Africa, 2014). However, since the introduction of antiretroviral therapy (ART), a significant decrease in mortality rate has been observed (Nolan *et al*., 2003).

The first antiretroviral (ARV) drugs to exhibit clinical efficacy for management of HIV infection were the nucleoside analogue reverse transcriptase inhibitors (NRTIs). The mode of action of NRTIs is to inhibit the HIV reverse transcriptase from converting viral RNA to DNA (Stankov *et al.,* 2010; Nolan *et al.,* 2003). The NRTI drugs are nucleoside analogues that act as competitive substrates along with natural endogenous dNTPs for binding by HIV reverse transcriptase, and cause truncation of viral DNA chain elongation when incorporated (Nolan *et al.,* 2003). The NRTIs include drugs such as zidovudine (AZT), lamivudine (3TC), stavudine (d4T), didanosine (ddI), abacavir (ABC), tenofovir (TDF) and emtricitabine (FTC). In 1996, highly active antiretroviral therapy (HAART) was established (Lehman and Hoke, 2010). HAART consists of a combination of three or more antiretroviral drugs (Lehman and Hoke, 2010). HAART programmes ordinarily combine two NRTI drugs with drugs of other classes (using different mechanisms) such as the non-nucleoside reverse transcriptase inhibitors (NNRTI) or HIV protease inhibitors (PI) (Nolan *et al.,* 2003, Bindu and Anusha, 2011). The ability of HAART to eradicate the virus is impeded by the presence of cellular and anatomical reservoirs of HIV infection (Shen *et al.,* 2008), due to this reason HIV-positive individuals require a life time use of HAART in order to keep the viral load undetectable. In 2015 there were 36.7 million people living with HIV in the world and 17 million had access to ART (UNAIDS, 2015).

HAART has significantly decreased HIV mortality (Gonzalez-Duarte *et al.*, 2007); however, the rise in antiretroviral toxicity burden mediated specifically by NRTI therapy may offset these benefits (Nolan *et al.*, 2003). Despite the benefits of ART drugs, there are various long term complications associated with their use adverse effects accompanying HAART treatment have become increasingly common (Apostolova *et al.*, 2011b). In particular, the NRTIs are associated with mitochondrial toxicity which results in serious side effects such as peripheral neuropathy, lactic acidosis, hepatic steatosis, lipoatrophy and hyperlactemia (**Table 1.1**). Other adverse side effects such as neurological complications, dyslipidaemia, insulin resistance, dysglycemia, lipoatrophy, central fat accumulation have become evident in 40-60 percent of HIV-positive patients exposed to ART, especially d4T (Sinxadi *et al.*, 2013; Wiebe *et al.*, 2011).

Table 1.1. Adverse effects associated with NRTIs and the reported manifestations of mitochondrial toxicity (Apostolova *et al.*, 2011b).

Drug Family	Side Effects	Reported Manifestations of Mitochondrial Toxicity
NRTI	Hypersensitivity syndrome (fever, myalgia, headache, malaise, insomnia, nausea, vomiting, symptoms suggestive of upper respiratory tract infection, anorexia) Rash, Pancreatitis, Lactic acidosis Hepatotoxicity and hepatic steatosis Peripheral neuropathy Hyperpigmentation of palms and soles or fingers and nails Gastrointestinal effects (flatulence, nausea, diarrhoea, abdominal discomfort) Anemia, neutropenia, Myalgia, myopathy Dyslipidaemia and lipodystrophy	Inhibition of Pol G Depletion of mtDNA Reduction of mtDNA-encoded proteins Respiratory chain dysfunction Direct inhibition of ETC complexes (I, IV) Reduction of ATP levels ROS production Decrease in membrane potential Impairment of ADP/ATP translocase Impairment of fatty acid oxidation

In South Africa, HAART consists of standardised national ARV regimens, which include first, second and third line regimens (Nattrass, 2006; The South African Antiretroviral Treatment Guidelines, 2013). The South African public sector was

allowed access to ART drugs in 2004, and the first-line treatment in 2004 consisted of the antiretroviral drug stavudine (d4T) together with lamivudine (3TC) and an NNRTI either efavirenz (EFV) or nevirapine (NVP) (Nattrass, 2006; Orrell *et al.,* 2007). d4T was an affordable and efficacious ART drug and it was widely used as part of the first-line treatment of HIV in resource limited settings regardless of its detrimental effects on the neuronal cells (Phillips *et al.,* 2010). The second-line regimen consisted of lopinavir/ritonavir (LPV/r), AZT and didanosine (ddI), and it was offered to patients on whom the first line treatment failed (Orrell *et al.,* 2007).

However due to the high toxicity of d4T, The World Health Organization (WHO) has since 2010 recommended the use of TDF instead of stavudine in national programs (Brennan *et al.,* 2012). The need for better treatment outcomes through reduced toxicity and better ARV drug adherence was the reason for the change (Brennan *et al.,* 2012). Currently (i.e. post 2010) the first-line regimen in South African guidelines consists of a fixed-dose combination (FDC) that contains 300 mg TDF, 200 mg FTC or 3TC and 600 mg efavirenz (EFV) (Davies *et al.,* 2013).

1.2 Sensory neuropathy

Sensory neuropathy (SN), or peripheral neuropathy, is a frequent side effect of HIV disease and perturbs the life of patients infected with the virus (Wadley *et al.,* 2011). SN is a common neurological problem in HIV-positive individuals both on and off ART. A painful sensory neuropathy is often developed by about 30% to 60% of patients harbouring the HIV virus (on and off ART) and it is usually characterised by degeneration of both unmyelinated and myelinated nerve fibres in a length dependent manner (Kamerman *et al.,* 2012; Keswani *et al.,* 2002). Previous studies in South Africa reported that about 60% of HIV-positive outpatients receiving ART treatment were diagnosed with SN (Maritz *et al.,* 2010; Wadley *et al.,* 2011; Wadley *et al.,* 2014).

SN experienced in HIV infection comprises of two types that are clinically undistinguishable, namely:

- Distal symmetric polyneuropathy (DSP) which develops due to the HIV virus itself and is usually a complication of advanced immunosuppression by HIV infection (Phillips *et al.,* 2010). Neurons do not express CD4+ receptors, thus neuronal damage observed in HIV-positive patients its unlikely to have been due to active infection of neurons (Acharjee *et al.*, 2010), but may arise due to infiltration of activated immune cells into the peripheral nerves.

- Antiretroviral toxic neuropathy (ATN) which results due to toxicity induced by exposure to NRTI antiretroviral treatment (Hahn & Husstedt, 2008). Subsequent to the introduction of ART, an increase in the incidence and prevalence of SN has been observed (Wadley *et al.,* 2011). This escalation is associated with the use of ART treatment including NRTI such as d4T (Wadley *et al.,* 2011).

SN is a disease that is characterised by damage to distal peripheral nerves (i.e. peripheral neuropathy) approximately within the same areas on both sides of the body. It typically starts by affecting hands and feet, which may further advance to the arms and legs. Degeneration usually affects the distal portions of axons first and atrophy of axons progresses gradually in the direction of the nerve cell body (Hughes, 2002). Pain is the most common feature of SN, and it is usually bilateral, of slow onset and often described as "painful numbness, aching, or burning sensation" (Hao, 2013). Other signs and symptoms include a sense of pins and-needles, loss/reduced reflexes of ankles and loss of sense of vibration (Wadley *et al.*, 2011).

The most important mediators of SN are the HIV virus, HAART drugs especially NRTIs, and the immune response to the HIV virus. These are all, either in combination or individually potentially toxic to the nervous tissues (Kamerman *et al.*, 2012). An improved understanding of how these three factors bring about the clinical phenotype is limited which has resulted in a delay for the development of efficacious therapeutic intervention (Maritz *et al.,* 2010).

The histological feature that seems to be common in patients with both DSP and ATN is severe loss of dorsal root ganglion (DRG) sensory neurons, DRG infiltration by macrophages infected with HIV, degeneration of distal long axons and a "dying back" neuropathy (Hao, 2013). Degeneration of axons and the reduction of nerve fibre density across the nerve trunk periphery, neuronal loss in the DRG and a dying-back of peripheral nerves are characteristics of SN. Nerve biopsies obtained from individuals who died due to the advancement of the disease, have shown the presence of immune cell infiltration and accumulation of inflammatory markers released into DRG and peripheral nerves (Verma and Simpson, 2007).

The use of immunohistochemistry technique reveals macrophage activation within the epineurium of peripheral and dorsal nerve ganglia (Verma and Simpson, 2007). These reactive macrophages in regions of axonal degeneration cause the local release of pro-inflammatory neurotoxic cytokines with expression of TNF-alpha, interleukin IL-1 and IL-6 (Hao, 2013; Wadley *et al.*, 2013). This cytokine response is essential for the removal of the pathogen (HIV); but a prolonged or large response can result in pathology such as SN (Verma and Simpson, 2007). An increase above normal levels of TNF-alpha in HIV-1 positive individuals might also lead to extra neurodegeneration (Hao, 2013).

A better understanding of the pathogenesis of SN is needed in order to better understand and identify predisposing factors which will aid in the implementation of effective preventative and pain management strategies (Phillips *et al.*, 2010).

1.3 Mitochondrial toxicity of ART (*POLG* hypothesis)

In healthy individuals, mammalian somatic cells contain up to several thousand mitochondria, with each mitochondrion containing 1–10 copies of mitochondrial DNA (mtDNA) (Cummins, 1998). Therefore, normal human cells contain ~1 000 to 10 000 copies of their mtDNA genome per cell. This is not a fixed number and the tissues' energy demand has a regulatory effect on mtDNA copy number (Lee & Wei, 2005, Shen *et al.*, 2008). mtDNA copy number can be described as an absolute number, or can be described in comparison to the amount of nuclear

DNA (nDNA) present i.e. the mtDNA:nDNA ratio or relative mtDNA copy number.

It has been shown that HIV infection decreases mtDNA copy number. A study by Côté *et al.* (2002) indicated that HIV infected patients (ART-naïve) had reduced levels of mtDNA compared to HIV-negative and age matched patients used as controls. HIV-positive patients were found to have approximately 34% reduction in mtDNA: nDNA ratio as compared with HIV-negative controls in peripheral blood mononuclear cells (PBMC). HIV infection is able to impede the normal physiological function of the mitochondria in lymphocytes, inducing mtDNA loss.

ARV use can further decrease mtDNA copy number (refs). In Cote *et al.* (2002), patients in receipt of d4T-containing ART who developed symptomatic lactic acidemia had even further reduced level of mtDNA: nDNA ratio (68% of levels in HIV-negative controls) even lower than in the ART-naïve HIV-positive cohort (Côté *et al.* 2002). There was a significant increase in mtDNA levels once ART was stopped (**Figure 1.1**).

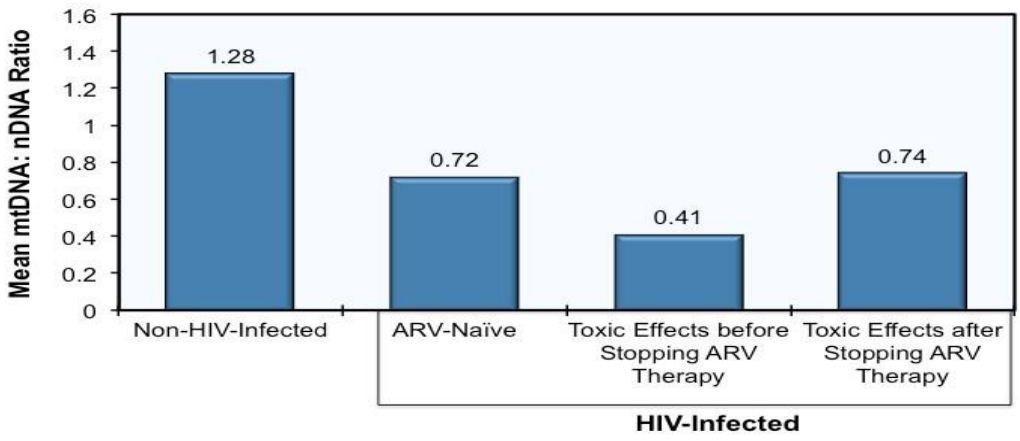


Figure 1.1. Average mtDNA: nDNA ratios in peripheral blood samples obtained from different groups of HIV-infected patients. (Côté *et al.*, (2002); http://depts.washington.edu/hivaids/arvae/case5/discussion.html).

de Mendoza *et al.*, (2004) also indicated that d4T induced mtDNA depletion in PBMC from HIV-positive patients on ART as compared to HIV-positive treatment naïve and HIV-negative individuals (**Figure 1.2**). de Mendoza *et al.*,

(2004) concluded that HIV on its own and together with d4T treatment can induce mtDNA copy number loss in PBMCs but other types of HAART were not as toxic. These results were similar to those in a clinical trial by Gallant *et al.* (2004), whereby TDF was compared with d4T and TDF was found to be less toxic (*in vitro*) to the mitochondria than d4T.

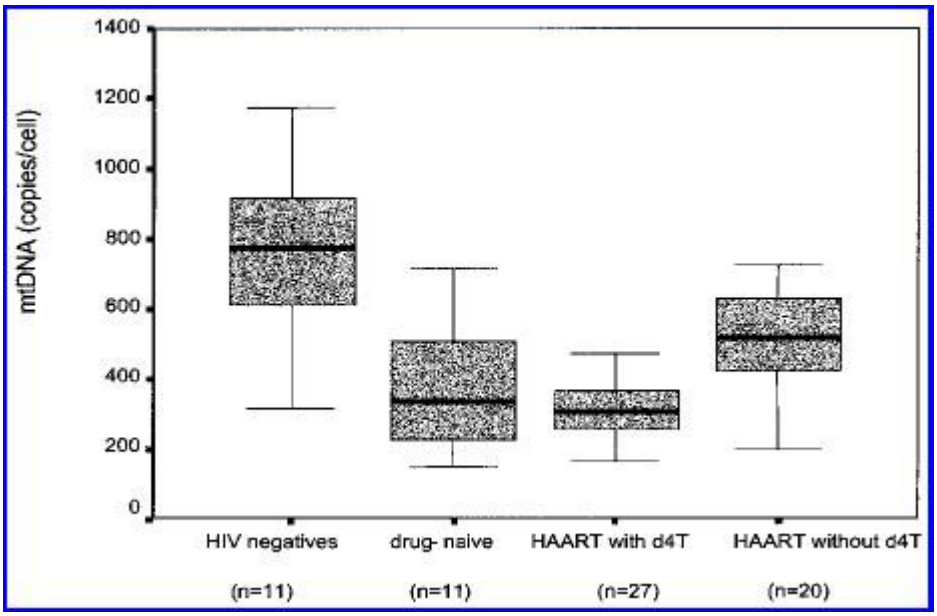


Figure 1.2. mtDNA copy number in HIV-negative patients vs. HIV-positive patients who were either drug-naïve or using d4T or other ART (from de Mendoza *et al.*, 2004).

Menezes *et al.* (2013) investigated the early effects of d4T in comparison with TDF on mtDNA copy number in adipose tissue in black South African patients. The study cohort (n = 60) consisted of HIV-positive patients taking varying doses of d4T or TDF based ART. The patients who were exposed to the standard and low dose d4T had a decrease in mean mtDNA copy number/cell of 29 % and 32 %, respectively when compared with TDF after only four weeks.

The mechanism by which NRTIs decrease mtDNA copy number is thought to be via the "*POLG* hypothesis" (Brinkman 1999). NRTIs inhibit the reverse transcriptase enzyme of HIV by competing with endogenous nucleic acids for incorporation into nascent DNA. By the same mechanism, NRTIs also inhibit human DNA polymerase gamma (*POLG*) which is responsible for the replication of mtDNA. While there are several types of human DNA polymerase, mtDNA

POLG appears particularly susceptible to inhibition by NRTI, as it lacks the ability shared by nDNA polymerases to effectively discriminate against NRTI in favour of endogenous nucleic acids (Nolan *et al.*, 2003). NRTI inhibition of *POLG* leads to premature mtDNA termination, decreased mtDNA replication and decreased mtDNA copy number. In addition, the inhibition of *POLG* exonucleolytic proof reading capacity introduces errors during replication of mtDNA, and the usual mtDNA repair mechanism becomes less effective due to NRTIs ability to resist exonucleolytic removal (Lewis *et al.*, 2006; Brinkman *et al.*, 1998; Apostolova *et al.*, 2011a).

There are significant differences in the relative potencies of NRTI in their ability to interact with POLG and change mtDNA copy number. In particular, the NRTI "D–drugs" such as d4T, ddc, 3TC and ddl in their triphosphate state are more easily incorporated in the elongated strand of mtDNA *in vitro* than NRTIs such as ABC, TDF and AZT (Birkus *et al.*, 2002), therefore cause particularly lowered levels of mtDNA. 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)" (Birkus *et al.*, 2002, **Figure 1.3**).

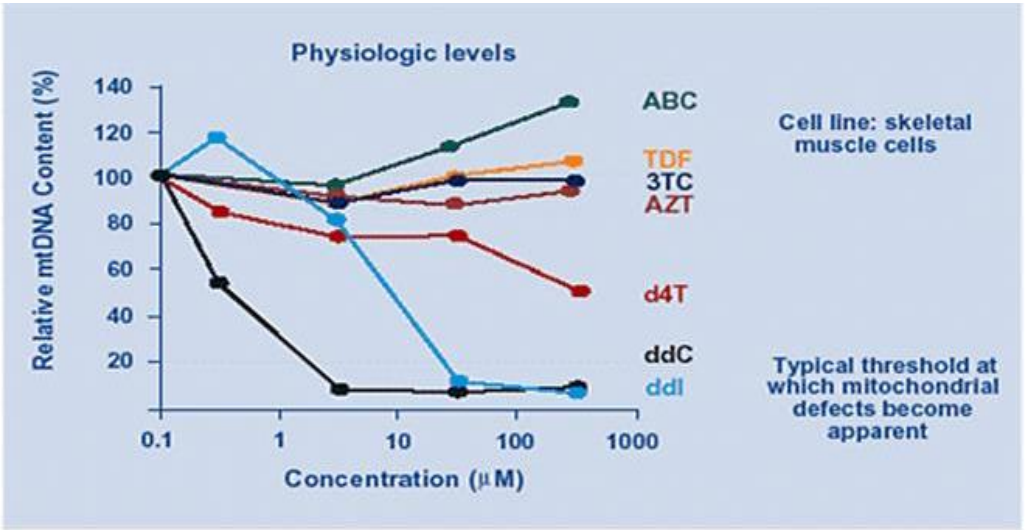


Figure 1.3. Depletion of mtDNA in skeletal muscle cells is particularly associated with D drug-NRTIs (d4T, ddC and ddI) (Birkus *et al.*, 2002)(http://www.medscape.org/viewarticle/466101).

1.4 Mitochondrial DNA copy number and ART side-effects

D-drug-dependent inhibition of mtDNA gamma polymerase and reduction in neuronal mitochondrial DNA copy number is thought to be the cause of SN in HIV infected individuals. NRTI use leads to reduced mtDNA copy number via inhibition of mtDNA synthesis. The enzyme polymerase gamma is inhibited by NRTI (**Figure 1.4**) which blocks replication of mtDNA, thus producing less adenosine triphosphate (ATP) and producing more reactive oxygen species (ROS) which damages the integrity of the mitochondrion and mtDNA (Hao, 2013). In a second mechanism, NRTIs are also thought to be able to act as stimuli for release of cytokines and chemokines by glia, which further induces neuronal sensitisation (Hao, 2013).

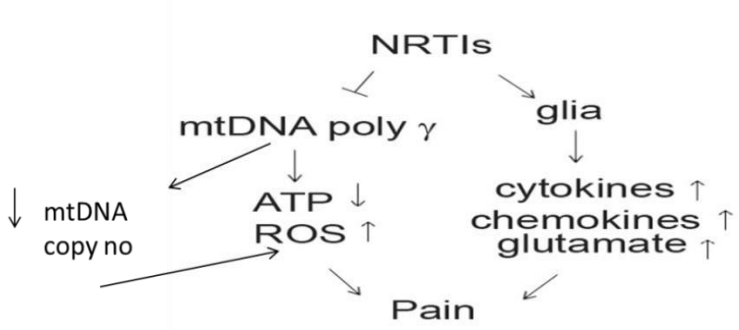


Figure 1.4. A model of two ways in which NRTIs may induce neuropathic pain in HIV patients on ART (modified from Hao, 2013).

Several studies have suggested that changes in mtDNA copy number may be linked with occurrence of ART-induced mitochondrial toxicities, including sensory neuropathy, lipodystrophy and changes in lactate levels. These are summarised below.

Dalakas *et al.* (2001) found that ddC use induced mitochondrial DNA loss in nerve cells of HIV-positive patients with Peripheral neuropathy. In order to understand the cause of the peripheral neuropathy, a molecular and morphological analysis was performed on nerve biopsy specimens from patients exposed to ddC (diagnosed with peripheral neuropathy) and from control subject with only peripheral neuropathy but not exposed to ddC. Abnormal and normal mitochondria were counted from the nerve cells, and mtDNA copy number

detection was conducted using qPCR. Morphological examination of the nerve biopsies indicated varying levels of axonal degeneration in all nerves. The ddC exposed patients showed severe myelin degeneration and 80% mtDNA copy number reduction in nerves as compared to the controls.

Kampira *et al*. (2014) investigated whether mtDNA depletion in whole blood could be used as a biomarker for d4T-induced mitochondrial toxicities in a case control study. A total of 203 HIV-positive patients exposed to d4T-containing ART from Malawi (n = 43 presented with SN, n = 20 presented with lipodystrophy, n = 113 presented with lactic acidosis) were compared to 64 healthy controls. Their results proved that mtDNA depletion from PBMCs can serve as a marker for lactic acidosis due to d4T use; however, did not support the use of mtDNA depletion as a marker for SN or lipodystrophy.

Chiappini *et al.,* (2004) found no association between mtDNA copy number and lipodystrophy or plasma triglyceride levels after AZT ART use. However, Chiappini *et al.,* (2009) explored the relationship between *POLG* gene SNPs and lipodystrophy due to d4T NRTI use. Three *POLG* SNPs (R1146, E1143 and E1146) and CAG repeats of *POLG* were studied in 69 HIV-positive patients treated with d4T and diagnosed with lipodystrophy compared to 138 patients without lipodystrophy. The E1143 SNP was significantly associated with both lipodystrophy and low mtDNA copy number in PBMCs after d4T ART use.

Many other studies outside the context of ART use have also shown that peripheral neuropathy is a common manifestation of mitochondrial disease or mitochondrial depletion. Luigetti *et al.* (2016) studied patients diagnosed with mitochondrial diseases such as Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE), Mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes (MELAS) and Myoclonic epilepsy with ragged-red fibres (MERRF). A total of 100 %, 92 % and 69 % of patients diagnosed with MNGIE, MELAS and MERRF presented with peripheral neuropathy, respectively.

In summary, several typical side-effects to NRTIs, including sensory neuropathy, lactic acidosis and lipodystrophy have been demonstrated to associate with low mtDNA copy number. Association between peripheral neuropathy and inherited mtDNA depletion syndromes has also been shown. These have been demonstrated in diverse tissues such as adipocytes, neurons, leukocytes and whole blood. However, these observations have not been consistently upheld, since some studies did not find significant associations between mtDNA depletion after ART use, and SN or lipodystrophy (Cherry *et al.,* 2002; Van Oosterhout *et al.,* 2013). Hence the use of mtDNA copy number as a marker for ART-induced toxicity remains debatable.

1.5 Pharmacogenetics of ART-induced sensory neuropathy

Pharmacogenetics is the study of genetic variants and their role in influencing inter-individual variability in responses to a pharmacologic therapy (Ortega & Meyers, 2014). Pharmacogenetics epitomizes a gene-environment interaction whereby genetic variation interacts with an exposure to the environment (drug) to modify a quantifiable phenotype associated to drug effectiveness or toxicity (Ortega & Meyers, 2014).

The fact that not everybody exposed to NRTIs develops SN, suggests that there may be an underlying genetic predisposition towards development of SN in response to ART use. Variation in several candidate genes has been suggested to modify the predisposition to developing SN (Hendry *et al.*, 2013). The pharmacogenetics underlying SN in the samples used in the current study have already been studied extensively and the following genes (**Table 1.2**) have been examined so far.

A genome wide association (GWAS) study of peripheral neuropathy with D-drug containing regimens in HIV-positive patients was performed by Leger *et al.*, (2014). A total of 254 participants including 90 SN cases and 164 controls without SN from the AIDS Clinical Trials Group protocol 384, using d4T or ddl-containing ART, were studied. Approximately one million genome-wide SNPs, as well as 1 395 selected SNPs in 46 candidate genes associated with peripheral neuropathy, were investigated for association with SN. The cohort was made up of Black (34%), Hispanic (17%) and Caucasian (49%) individuals. No SNPs were consistently found to be associated with new onset peripheral neuropathy at genome wide significance. Susceptibility to d4T/ddI-associated neuropathy was not explained by a single genetic variant with a marked effect.

However, the following SNPs *CCDC19* rs2501325, *LITAF* rs13333308, *ZNF648* rs7554182, *NFE2L3 (NRF3)* rs4722585 from Leger et al. (2014) were chosen for replication in our study due to their low P-values and high association odds ratio (OR) with peripheral neuropathy at genome wide level. *ZNF648* rs7554182 had the lowest GWAS P-value ($P = 5.7 \times 10^{-08}$, OR = 48.5) in Black individuals with

grade 3 SN only, defined as sensory alteration or paraesthesia causing inability to perform usual social and functional activities. *NFE2L3 (NRF3)* rs4722585 had the lowest GWAS P-value ($P = 4.4 \times 10^{-06}$, OR = 0.2) in Black individuals, in association with grade 2 SN in Leger et al. (2014), where grade 2 was defined as sensory alteration or paraesthesia causing greater than minimal interference with usual social and functional activities. *CCDC19* SNP rs2501325 and *LITAF* SNP rs13333308 had lowest P-values ($P = 1.3 \times 10^{-03}$, OR = 0.2 and $P = 6.0 \times 10^{-06}$, OR = 32.3) in analysis using 46 candidate genes only in Blacks with grade 2 and grade 3 SN respectively. Leger *et al.*, (2014) suggested that these results be replicated in other cohorts of HIV-positive patients diagnosed with SN.

Table 1.2. Genes already examined in association with SN in various cohorts.

Examined Genes	Cohort(s)	Reference(s)	Phenotype examined	Study conclusions
POLG	Malawians, French	Van Oosterhout Chiappini	Peripheral Neuropathy, Lipodystrophy, mtDNA copy number	None Significant
MHC III (BAT & TNFA)	Africans, Chinese, Indonesians, Malaysians & Caucasians. South Africans (current study cohort)	(Cherry *et al.*, 2008), (Chew *et al.*, 2011), (Affandi *et al.*, 2008) (Wadley *et al.*, 2014)	Lipodystrophy, Pain Intensity, Peripheral Neuropathy	Significant Association
Cytokine s	Africans & Europeans	(Cherry *et al.*, 2008)	Pain Intensity Peripheral Neuropathy and	Significant Association
Cytokines	Southern Africans	Wadley *et al.*, 2013	Sensory Neuropathy, Pain Intensity	Significant Association
Mitochondrial uncoupling Proteins (UCP)	South Africans (current study cohort)	(Wadley *et al.*, 2013)	Pain Intensity Peripheral Neuropathy and	None Significant
GTP cyclohydrolase 1 (GCH1)	South Africans (current study cohort) & Europeans	(Wadley *et al.*, 2012)	Pain Intensity Peripheral Neuropathy and	None Significant
KCSN1	South Africans (current study cohort)	(Hendry *et al.*, 2013)	Pain Intensity	Significant Association
mtDNA haplogroups	Africans (Black Malawians) South Africans (current study cohort)	(Kampira *et al.*, 2013); (Goldfein, 2014)	Lactic Acidosis and Peripheral Neuropathy; CD4 count	None Significant
GWAS	Americans (Blacks, Hispanic, and Caucasians)	Leger *et al.*, 2014	Sensory Neuropathy	None significant at genome-wide level

1.6 Genetics of mitochondrial DNA copy number

In the current pharmacogenetic study, we have chosen to study mtDNA copy number and genetic variation in nuclear genes which influence mtDNA copy number. Variation in genes controlling mtDNA copy number may influence susceptibility to NRTI-induced mtDNA copy number- mediated side effects. Host genetic variation of the mtDNA and/or nuclear-encoded mtDNA genes may play a vital role in determining the degree of d4T toxicity (Sinxadi *et al.,* 2013) and mtDNA depletion (Zhang & Singh, 2014). Genes that regulate mitochondrial copy number may harbour genetic variation that may cause an underlying susceptibility to mtDNA copy number depletion that is not phenotypically noticeable until NRTIs are used. There is a concept of a threshold level of mtDNA depletion before symptoms are noticeable. The mtDNA copy number may be lowered to a subclinical level due to inherited genetic mutations in genes controlling mtDNA copy number, which then get further lowered past the threshold point by ART, causing symptoms (Montier *et al.,* 2009).

Yamanaka *et al*. (2007) and Bailey *et al*. (2009) suggested that NRTI treatment in the context of *POLG* mutation may reduce mitochondrial function below clinical threshold, causing mitochondrial disease. A common SNP in *POLG* is E1143G found in about 4% of the general population and associated with functional impairment of *POLG* (Copeland, 2010). Hudson *et al*. (2006) observed that functional genetic variants of *POLG* are present in up to 0.5% of the general population. It is possible that there are genetic polymorphisms present at reasonable frequency in South African populations within *POLG* or other candidate genes that may make them more susceptible to the effects of d4T.

POLG, the gene encoding polymerase gamma enzyme has been the major nuclear gene that has previously been studied in ART-induced SN and ART-induced changes in copy number. Mutations in *POLG* have been comprehensively studied for their role in mitochondrial copy number disorders (Ponamarev *et al.,* 2002. Van Goethem *et al.,* 2001. Longley *et al.,* 2005). The National Institute of Environmental Health Sciences (NIEHS) has developed a *POLG* database that catalogues more than 120 *POLG* mutations and their associated disorders

(http://tools.niehs.nih.gov/polg/). Yamanaka *et al.* (2007) reported that a novel homozygous DNA *POLG* mutation (R964C) was associated with increased mitochondrial toxicity from NRTI treatment in a Japanese woman with HIV/AIDS. Yamanaka *et al* (2007) using cell lines treated with NRTI d4T, found that cells harbouring the *POLG* mutant R964C had considerably decreased mtDNA levels as compared to the wild type *POLG*. In patients diagnosed with d4T induced lipodystrophy the E1143G mutation was significantly associated with reduced mtDNA content (Chiappini *et al.,* 2009). Chen *et al.,* (2002) found no association between *POLG* CAG repeat length in the second exon and lactic acidosis or neuropathy in HIV-positive Dutch patients exposed to ART. In a cohort study of HIV-positive patients from Malawi treated with d4T containing ART and experiencing severe d4T induced side effects, *POLG* was sequenced, but it was found that no mutations were harboured by the patients (Van Oosterhout *et al.,* 2013; Kampira *et al.,* 2013). This might suggest that monogenic *POLG* mutations are not a common pathogenic determinant of severe d4T-associated mitochondrial toxicity in Malawians (Van Oosterhout *et al.,* 2013; Kampira, 2013).

To identify genetic variation in other nuclear genes apart from POLG which influence mtDNA copy number, an understanding of mtDNA replication is required. A brief description of mtDNA replication and the genes involved follows, with a suggestion of candidate genes to be considered in this study.

The mitochondria are cellular organelles that are ubiquitous in eukaryotes and are sites of vital cellular functions (Montier *et al.,* 2009) such as programmed cell death, cellular proliferation, signal transduction (Zhang and Singh, 2014), synthesis of key catalytic subunits of the electron transport chain complexes and synthesis of ATP (Suomalainen and Isohanni, 2010). The primary function of the mitochondria is the production of ATP molecules through the process of oxidative phosphorylation, which serves as energy for cellular functions (Montier *et al.,* 2009).

Human mitochondria contain numerous copies of double stranded haploid DNA of approximately 16 kb circular genome which encodes 13 proteins responsible for

proper functioning of electron transport chain, as well as 22 tRNAs and 2 rRNAs needed for the translation of the encoded proteins within the mitochondria (Scarpulla, 2008). The 37 mitochondrial genes are encoded by both of the two strands of the double stranded DNA designated heavy and light (Arnold *et al.*, 2012). Human mitochondria also contain a displacement loop (D-loop) or control region which is the main non-coding region of human mtDNA and it includes promoters which instigate transcription (Arnold *et al.,* 2012). Heavy and light strands contain two heavy-strand promoters and one light-strand promoter respectively (Falkenberg *et al.,* 2007).

mtDNA replication is accomplished by proteins encoded by genes from the nuclear genome (Zhang and Singh, 2014; Montier *et al.,* 2009), these proteins translocate to the mitochondria. Unlike nDNA replication which is dependent on cell division, mtDNA replication is independent of the cell cycle (Hudson and Chinnery, 2006). Mitochondrial replication is specifically suited to make as many mitochondria as that particular cell needs at the time. There are two suggested models of mtDNA replication, namely, the simultaneous and the asynchronous model (Carling *et al.,* 2011). Both models have delayed synthesis of the DNA-lagging strand.

Replication and transcription of mtDNA require RNA primers. The RNA primer needed for first strand mtDNA replication is synthesised by the mitochondrial RNA polymerase (POLRMT) with the assistance of mitochondrial transcription factors A, B1 and B2 (TFAM, TFB1M and TFB2M respectively) (Chang and Clayton 1985; Fish *et al.,* 2004). TFAM is a key regulator in the activation of replication and transcription of mtDNA which bends mitochondrial promoter DNA to aid transcription of the mitochondrial genome. TFAM expression is influenced by other transcription factors such as nuclear respiratory factor 1 (NRF1), nuclear respiratory factor 2 (NRF2) and peroxisome proliferator-activated receptor co-activator 1 alpha (PGC-1a). PGC-1a is another key element regulating mitochondrial biogenesis. PGC-1a is also known to activate the expression of NRF1 and sometimes NRF2 (Austin & St-Pierre, 2012).

NRF2 is a transcription factor that regulates the expression of antioxidant genes by binding to their promoter NRF2-antioxidant response element (ARE) (Shimoyama *et al.,* 2014). NRF2 is a key regulator of several pathways including mitochondrial biogenesis, mtDNA replication and maintenance, oxidative stress (Nagiah *et al.,* 2015). Both HIV and ARTs increase oxidative stress (Reddy *et al.,* 2012, Nagiah *et al.,* 2015) which may increase NRF2 expression (Reddy et al., 2012). Proteins involved in mtDNA transcription that also help to initiate DNA replication are summarised in **Table 1.3** below.

Table 1.3. Proteins involved in mtDNA transcription that also help to initiate DNA replication

Protein	Gene	Chromosome	Function
Mitochondrial RNA polymerase	*POLRMT*	19	Synthesizes RNA primers for initiation of replication of the mitochondrial genome.
Mitochondrial transcription factor A (mtTFA)	*TFAM*	10	Key activator of mitochondrial transcription and participant in mitochondrial genome replication
Transcription factor B1	*TFB1*	6	Transcription Factor
Transcription factor B2	*TFB2*	1	Transcription Factor
Rnase MRP	*RMRP*	9	Initiates mitochondrial DNA replication
Endonuclease G	*ENDOG*	9	Modifies the RNA primers for mtDNA replication
Rnase HI	*RNASEH1*	2	Degrades the RNA primer hybridized to mtDNA.
Nuclear respiratory factor-1	*NRF1*	7	Transcription factor that binds to promoters of TFAM and POLG.
Nuclear respiratory factor-2	*NRF2*	2	Transcription factor that binds to promoters of TFAM and POLG.
PGC-1a	*PPARGC1A*	4	Transcription factor that binds to promoters of NRF1 and NRF2. Regulator of mitochondrial biogenesis and function.

Once initiated by the POLRMT complex, mtDNA replication is performed by the set of proteins, shown in **Table 1.4** and **Figure 1.5**. The enzyme POLG consists of one catalytic subunit which adds dNTPs to the newly synthesised DNA strand (p155) and two subunits that ensures high processivity of the catalytic subunit (p55) (Carling *et al.,* 2011). The subunits of the POLG enzyme are encoded by two genes, *POLG* and *POLG2* respectively. *POLG* activity is controlled by transcription factors NRF1 and NRF2. mtDNA replication is further supported by Twinkle helicase (*C10orf2*) which unwinds short stretches of dsDNA in the 5′ to 3′ direction. mtSSBPs (mitochondrial single-stranded binding proteins) protect the single-stranded mtDNA (that serves as a template for lagging strand synthesis during mtDNA replication) from being digested by nucleases, and also prevent premature annealing or secondary structure formation (Dickinson *et al.*, 2013; Carling *et al.,* 2011). Mitochondrial topoisomerases and ligase III catalyse the transient breaking and re-joining of DNA strands.

Several studies have examined SNPs in the genes governing mtDNA replication in relationship to different phenotypes which may occur due to mtDNA depletion. For example, SNPs in these genes have been reported in association with Parkinson's disease, Alzheimer's Disease, Huntington's Disease, depression, cancer and others (see Table 3.5 in the Results section). There are very few studies of SNPs directly associated with mtDNA copy number variation in healthy individuals, although a few studies have examined SNP role in athletic ability.

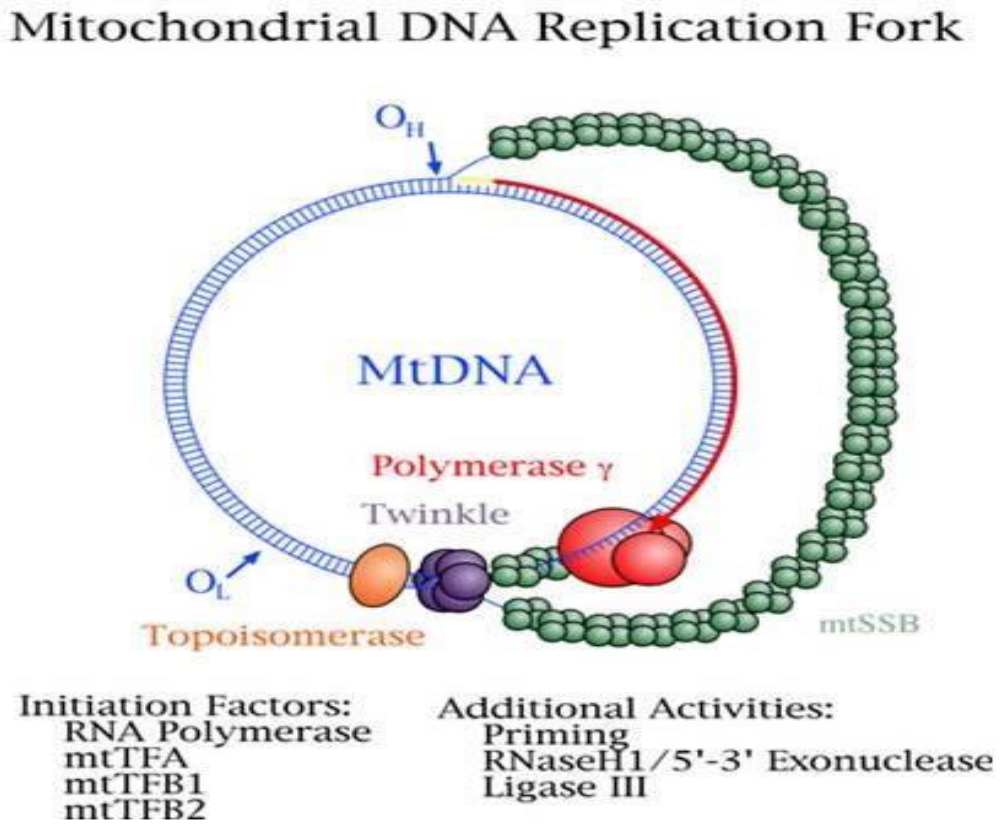


Figure 1.5: Mitochondrial DNA replication fork (Stumpf and Copeland, 2011). (https://www.niehs.nih.gov/research/atniehs/labs/gisbl/pi/mdnar/).

Table 1.4. Proteins involved in mtDNA replication. The genes listed here were the focus of this study.

Protein	Gene	Chromosome	Function
DNA polymerase gamma 1	***POLG***	15q25	mtDNA replication, builds new mtDNA nucleotide strand
DNA polymerase gamma 2	***POLG2***	17q23–24	Imparts high processivity to the catalytic subunit
Nuclear Respiratory Factor-2	***NRF2***	2	Transcription factor that binds to promoters of *TFAM* and *POLG*.
Twinkle Helicase	***C10orf2***	10q23.3	Unwinds duplex DNA
Mitochondrial single-stranded binding protein (mtSSB).	***SSBP1***	7q34	Prevent premature annealing, to protect ssDNA from being digested by nucleases.
Mitochondrial Topoisomerase 1	***TOP1***	10q12-13.1	Catalyses the transient breaking and re-joining of a single strand of DNA
Topoisomerase IIIa	***TOP3a***	17p11-11.2	Controls and alters the topologic states of DNA
Mitochondrial Transcription Factor A	***TFAM***	10	Key activator of mitochondrial transcription and participant in mitochondrial genome replication
DNA ligase III	***LIG3***	17q11.2-12	Ligates the DNA strands

Further potential candidate genes influencing mtDNA copy number come from studies of mtDNA depletion syndromes (MDS), which are a group of disorders inherited in an autosomal recessive pattern, that have severe decrease in mtDNA copy number in affected tissues and organs (El Hattab, 2013). Mutations in nuclear genes involved in pathways responsible for mtDNA maintenance, mitochondrial (dNTP) synthesis and mtDNA replication (such as the genes shown in the Tables above) can cause MDS. Genetic variants in proteins involved in the synthesis or regulation of mtDNA replication and dNTP pools have been associated with severe reduction in mtDNA copy number, which if reduced below 30% of normal mtDNA content will result in severe and often lethal diseases (Cohen & Naviaux, 2010; Rötig and Poulton, 2009). Specifically, mutations in *POLG, C10orf2, MTSSB, TFAM, dGK, RRM2B, SUCLA2, TK2*, and *TYMP* are known to cause mtDNA depletion syndromes (Cohen & Naviaux, 2010).

PINK1 (PTEN Induced Putative Kinase 1) is another interesting gene that may influence mitochondrial copy number. Whilst it is not directly involved in mtDNA transcription or replication, it plays a role in mitochondrial quality control, targeting destruction of dysfunctional depolarized mitochondria. The *PINK1* gene encodes for a protein kinase enzyme that localises to mitochondria (Gegg *et al.*, 2009). This enzyme targets faulty mitochondria for destruction since they lack sufficient membrane potential to allow translocation of PINK1 protein into the mitochondria, this will lead to excessive accumulation of the protein in the outer membrane. Mutations in *PINK1* gene are associated with Parkinson's Disease (PD), and mitochondrial involvement in the pathogenesis has also been observed. Silencing of *PINK1* is associated with mtDNA depletion (Gegg *et al.*, 2009).

In summary, we considered genes that play a role in mtDNA replication and/or transcription, and/or in mtDNA depletion syndromes, as candidate genes that may regulate mtDNA copy number and development of SN in response to d4T use.

AIMS AND OBJECTIVES

1.7 Hypotheses and Aims

In this study I examined variation in selected genes chosen from the Leger *et al.*, (2014) GWAS paper, as well as selected genes that maintain mtDNA integrity and replication. I hypothesized that:

1. Variation in candidate genes may associate with occurrence of SN after d4T-ART use
2. Variation in candidate genes may associate with pain intensity during occurrence of SN after d4T-ART use
3. Variation in candidate genes may associate with mtDNA copy number in patients using d4T-ART
4. Changes in mtDNA copy number may associate with occurrence of SN after d4T-ART use

The aims of this study were therefore to determine the associations between genetic variation in candidate nuclear genes, mtDNA copy number, occurrence of SN, and degree of pain intensity, in a Black South African cohort using d4T-ART. Given the d4T exposure, genotype-phenotype and phenotype-phenotype associations will be studied as shown in the following **Figure 1.6**.

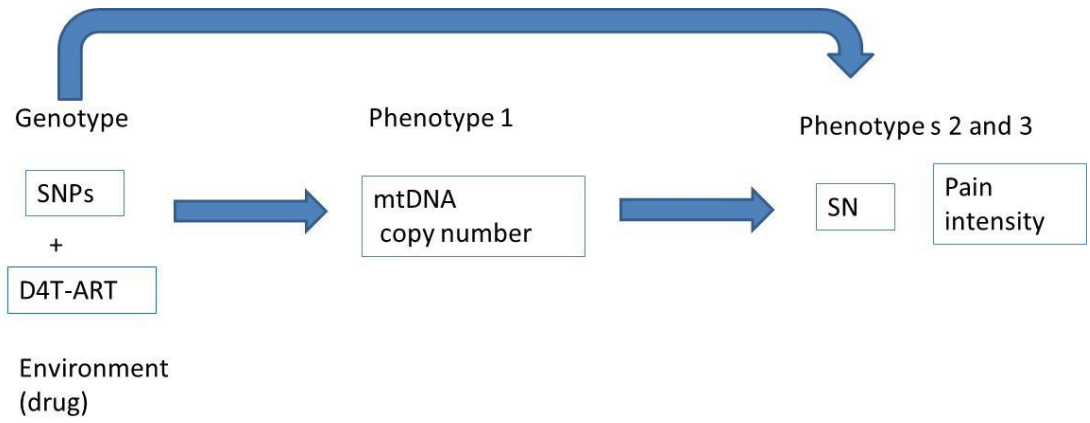


Figure 1.6. Genotype/phenotype associations and phenotype/phenotype associations (blue arrows) that will be examined in this study.

1.8 Objectives

1. To select approximately 30 candidate SNPs in nuclear genes that may influence mtDNA copy number, including genes driving mtDNA replication, and SNPs of interest from Leger *et al.,* (2014) paper.

2. To genotype candidate SNPs using MassArray technology, in a Black South African case-control cohort with and without SN after d4T use.

3. To examine statistical associations between candidate SNPs and SN phenotype in the Black South African case-control cohort, and to examine statistical associations between candidate SNPs and pain intensity in SN in the cohort.

4. To validate a qPCR assay to measure mtDNA:nDNA ratios (an indicator of mtDNA copy number).

5. To use qPCR to measure mtDNA:nDNA ratios in a Black South African case-control cohort with and without SN after d4T use.

6. To examine statistical associations between candidate SNPs and mtDNA: nDNA ratios in the cohort.

7. To examine statistical associations between mtDNA:nDNA ratios and SN in the cohort.

CHAPTER 2: MATERIALS AND METHODS

2.1 Case-control cohort

2.1.1 Sample collection

The samples used in this research project were obtained from a cohort of Black South Africans receiving HIV-treatment at Charlotte Maxeke Academic Hospital (Virology Clinic) Johannesburg, South Africa. Blood samples were collected by Dr Wadley (Brain Function Research Group, School of Physiology, University of the Witwatersrand) between July 2008 and April 2009 as part of her PhD study, approved by HREC (Medical), University of the Witwatersrand, protocol no. M080220 (**Appendix B**). The inclusion criteria for participation in the study were being HIV-positive, Black Southern Africans, with a minimum age of 18 years or older, and having received stavudine-ART for a minimum of six months prior to being screened for SN (Hendry *et al*., 2013; Wadley *et al.,* 2011). Written informed consent was obtained from all participants (Wadley *et al.,* 2013).

Ethical approval for the current genetics study was also obtained from the HREC (Medical) University of the Witwatersrand, protocol number M150459 (**Appendix A**). Of the original 342 samples analysed by Wadley *et al.,* (2013), a total of 272 samples were still available for use in the current study.

2.1.2 Phenotypic data collection and SN diagnosis

All participants were assessed by Dr Wadley at the time of blood sample collection. The following demographic and phenotypic data were collected: age, sex and ethnicity, height and weight; disease information (current and nadir CD4 T-cell counts, time since HIV diagnosis and AIDS-defining illnesses); ARV treatment history; and other possible causes of neuropathy (diabetes mellitus, alcoholism, vitamin B12 deficiency and exposure to isoniazid and chemotherapy) (Wadley *et al.,* 2011).

The patients were screened for sensory neuropathy with the aid of the validated AIDS Clinical Trials Group SN screening tool (Cherry *et al.,* 2005). The following subjective symptoms were recorded: pain, aching, burning, numbness and a sense of pins-and needles. It was also recorded whether SN symptoms were

experienced currently (on day of assessment) or ever (in past but not on day of assessment). Objective signs including absent ankle reflex and/or reduced sense of vibration were measured by Dr Wadley on the day of assessment. Vibration sense was assessed using a 128Hz tuning fork, which was placed on the interphalangeal joint of each great toe; vibration sense of ten seconds or less was considered abnormal. The combination of at least one (subjective) symptom and at least one (objective) sign was indicative of SN presence (Wadley *et al.,* 2011). Pain intensity was documented in patients with pain and was rated on a scale of 0-10.

2.1.3 DNA isolation

DNA was extracted from blood samples from the cohort using a salting out method (Miller, 1988) by Dr Wadley and stored at -20 °C in nuclease free water (Wadley, 2013). For the current study, stored DNA from the remaining 272 samples was assessed using a NanoDrop® ND-1000 Spectrophotometer, Absorbance values at 260nm and A280nm were measured and DNA concentration and quality were calculated. A total of 263 samples had DNA of acceptable quantity and quality (A260:A280 ratio ~1.8) for the current study. The stock DNA of all 263 samples was further diluted to a working concentration of 25 ng/µl for high throughput genotyping and for mtDNA quantification.

2.2 mtDNA copy number assessment

We **c**ompared the levels of nuclear to mitochondrial DNA (mtDNA) in human DNA samples using quantitative PCR (qPCR). qPCR was conducted with a Light Cycler® 480 qPCR instrument (Roche) and SYBR Green chemistry. SYBR® Green I (Thermo Fisher Scientific) is an affordable master mix for use as a mode of detection of DNA when using qPCR Instrument. SYBR Green is a fluorescent intercalating dye which binds to the double stranded DNA (**Figure 2.1**) and emits a fluorescent signal upon binding. In qPCR, DNA accumulates and fluorescent signal increases proportionally to the DNA concentration. The excitation and emission of maxima SYBR® Green I are at 494 nm and 521 nm, respectively, which are compatible with the use on any real-time cycler.

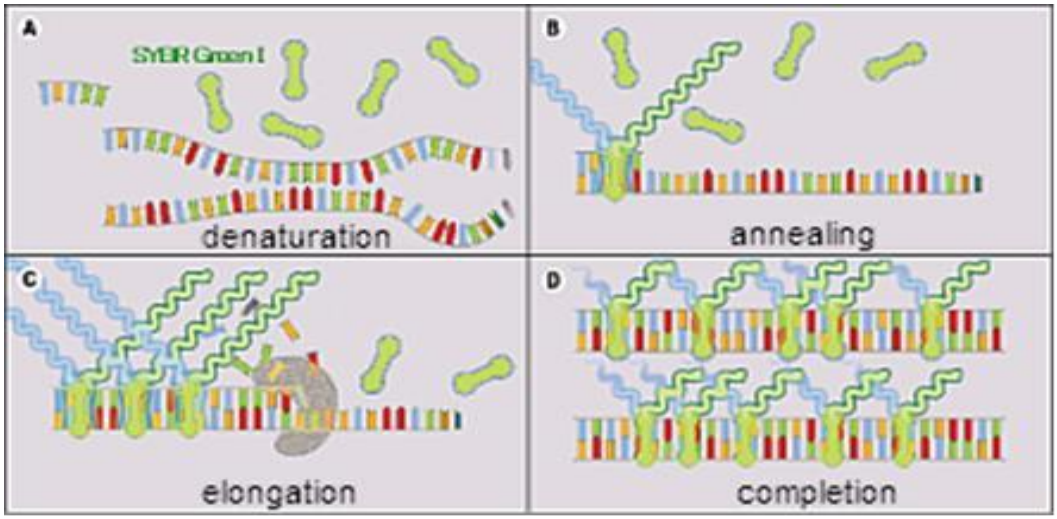


Figure 2.1: qPCR quantification of DNA using SYBR Green I for detection (Ponchel *et al.*, 2003) (http://www.ngrl-japan.com/gene/gene.html).

(A) The dye does not bind to ssDNA during denaturation.

(B) Once annealing is initiated the dye molecules will start binding to the dsDNA.

(C) During elongation, the dye attached to the dsDNA will emit fluorescence, (D) Once the amplicon is fully elongated, the fluorescence intensity will increase and will be detected by the qPCR instrument.

"Primer-dimers" or non-specific PCR products also emit fluorescence since they are also double stranded DNA. This is a major disadvantage of the SYBR Green chemistry and this can give false positives of the results. The primer-dimer issue can be avoided when designing primers and by reducing the concentration used. In addition, specificity of the PCR reaction must be confirmed as non-specific PCR products could contribute to SYBR Green fluorescence. Measuring fluorescence of double stranded mtDNA in comparison to fluorescence of double stranded nDNA allows the calculation of relative mtDNA copy number.

2.2.1 PCR Optimisation

Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) and nicotinamide adenine dinucleotide (*NADH1/MT-ND1*) genes were chosen as representatives of the nDNA and mtDNA respectively. *GAPDH* is known to be present as a single copy in the nuclear genome (Atema *et al.*, 2013) and no pseudogenes of *NADH1* are present in the nuclear genome (Grady *et al.*, 2014). These genes have been used

successfully elsewhere in mtDNA copy number estimation (Côté *et al.,* 2002; Rooney *et al.,* 2015). We performed qPCR using DNA and not mRNA, thus a single reference gene was deemed sufficient since we were not measuring gene expression (Kampira *et al.,* 2014; Grady *et al.,* 2014). We used two sets of PCR primers from each of these genes as shown in **Table 2.1** and the primer pair with highest specificity were chosen for use in the current study.

Table 2.1. nDNA & mtDNA qPCR primers and PCR conditions.

	Genes	Reference	Forward primer	Reverse Primer	Annealing temperature	Amplicon Length
nDNA	*GAPDH*	Primer Pair 1(Cooray *et al.,* 2002)	5'-TGACAACGAATTTGGCTACA-3'	5'-GGGGTCTACATGGCAACT-3'	55	196 bp
		Primer Pair 2 (Toyama *et al.,* 2003)	5'-AAATCAAGTGGGGCGATGCTG-3'	5-'GCAGAGATGATGACCCTTTTG-3'	60	208 bp
mtDNA	*NADH1*	Primer pair 1 (Psarra and Sekeris, 2011)	5-'ATGGCCAACCTCCTACTCCTCATT-3'	5'-TTATGGCGTCAGCGAAGGGTTGTA-3'	60	154 bp
		Primer pair 2 (Bonod-Bidaud *et al.,* 2001)	5-'CTAGCCCCCATCTCAATCATA-3'	5'-GAATGCGGTAGTAGTTAGGAT-3'	55	243 bp

Both PCR primer pairs were used during optimisation assays for each gene using regular PCR and qPCR. First we optimised the primers and PCR reactions by using conventional endpoint PCR and gel electrophoresis. The PCR master mix (KapaTaq) from Lasec and its PCR reaction mix (**Table 2.2**) were used to prepare a 20 µl reaction mix. PCR was done initially with MJ Mini™ Thermal Cycler (Gradient PCR) from Bio-Rad and thermal cycling conditions shown in **Table 2.3**). Annealing temperatures in **Table 2.4** were varied during optimisation with the standard PCR.

The amplicons were run on a 1% agarose gel using 1X TBE buffer solution from Life Technologies™ for 40 minutes at 80V. Five µl of amplicons was loaded per well using 5 µl of 1X loading dye. The loading dye was supplied by Applied Biosystems and gel-staining GR-green was supplied by Inqaba Biotec. The gel image was captured with the aid of the Gel-Doc visualisation instrument from Bio-Rad. The best annealing temperature was chosen for both target and reference genes. Then using the same conditions, the primer specificity was tested on a qPCR instrument.

Table 2.2. KapaTaq (20 µl) regular PCR reaction mix.

Reagent	**Volume**
KapaTaq (2X) Master Mix	10.0 µl
Forward Primer (10 µM)	0.6 µl
Reverse Primer (10 µM)	0.6 µl
Template DNA (25 ng/µl)	3.4 µl
Water, Nuclease-Free	5.4 µl
Total Volume	20 µl

Table 2.3. KapaTaq cycling conditions.

Step	Temperature (°C)	Duration	Cycles
Initial denaturation	95	3 Minutes	1
Denaturation	95	30 Seconds	35
Annealing	T_m -5	30 Seconds	
Extension	72	1 Minute	
Final Extension	72	1 Minute	1
Hold	4-10	∞	1

Next the qPCR assay was optimised on a Light Cycler 480 qPCR instrument by replicating the conditions used during the endpoint PCR optimisation. SYBR Green master mix from Thermo Fisher Scientific was used for detection. Thermal cycling conditions in **Table 2.5** and the qPCR reaction mix in **Table 2.6** were used for qPCR optimisation. Standard curves were generated for the calculation of PCR reaction efficiency. Triplicates of the following DNA concentration (i.e. 20 ng/µl; 2 ng/µl; 0.2 ng/µl and 0.02 ng/µl) were used for the standard curves, plus a negative control. qPCR reaction efficiency was calculated with LightCycler 480 SW 1.5 software. Melting curves (**Table 2.5** for thermal conditions) were also generated for measure of primer specificity. Generally, a qPCR efficiency between 90 and 110% was considered acceptable (Taylor *et al.,* 2010; Pettengil *et al.,* 2012).

Table 2.4. Annealing Temperature ranges used for PCR optimisation.

	Genes	Primer Set	Annealing Temperature Ranges (°C)	Best Annealing Temperature (°C)	Amplicon Length
nDNA	*GAPDH*	1(Cooray *et al.*, 2002)	48-53	50	196 bp
		2 (Toyama *et al.*, 2003)	47-58	57	208 bp
mtDNA	*NADH1*	1 (Psarra and Sekeris, 2011)	48-59	58	154 bp
		2 (Bonod-Bidaud *et al.*, 2001)	45-50	50	243 bp

The following qPCR conditions (**Table 2.5**) and qPCR recipe (**Table 2.6**) were chosen as optimal. White 96-well plates (Bio-Rad) compatible with the Light Cycler 480 qPCR instrument were used.

Table 2.5. Thermal cycling conditions as recommended by the SYBR Green master mix protocol.

Step	Temperature (°C)	Time	Number of Cycles
Initial Denaturation	95	10 Minutes	1
Denaturation	95	15 seconds	40
Annealing	*GAPDH* Pair 1 (50) *NADH1* Pair 1 (58)	30 seconds	
Extension	72	30 seconds	
Melting Curve	97	1 Minute	1
	40	10 Seconds	
	95	Continuous	
Cooling	40	40 Seconds	1

Table 2.6. qPCR recipe utilised in this study for quantification of mtDNA.

Reagent	Volume
Maxima SYBR Green (Thermo Fisher)	12.5 µl
Forward Primer (10 µM)	0.6 µl
Reverse Primer (10 µM)	0.6 µl
Template DNA (25 ng/µl)	1 µl
Water, Nuclease-Free	10.3 µl
Total Volume	25 µl

2.2.2 qPCR assay validation

We validated our qPCR assay using cell lines depleted of mtDNA. Treatment of eukaryotic cells with ethidium bromide (EtBr) depletes endogenous mtDNA while nDNA is maintained (Seidel-Rogol and Shadel, 2002). The loss of mtDNA extinguishes the oxidative phosphorylation system, after a few days, the cells completely devoid of endogenous mtDNA (ρ0 cells) die or need supplementation with nutrients to sustain viability (Kukat *et al.,* 2008). Treatment of cells with EtBr (50ng/ml) for more than 6 days results in extensive cell death (Seidel-Rogol and Shadel, 2002).

HEK293 cells (Sigma-Aldrich) were seeded in T-75 Corning® cell culture flasks, then grown in 80% HyClone™ DMEM medium (Thermo Fisher™), supplemented with 10% fetal bovine serum (FBS), 1% Pen-Strept (Penicillin + Streptomycin) and 1% Glutamine. The cells were cultured at 37 ℃ in a humidified incubator with 5% CO_2. Untreated cells were harvested by draining the media from the flask and washing the cells with 2 ml of Dulbecco's Phosphate buffered saline (PBS) (Thermo Fisher™). I then added 2 ml of 1X Trypsin EDTA (Lonza), and incubated the flask at 37 ℃ for two to three minutes; then 1 ml of cells was transferred into a new flask. This was done for each harvest time point. DMEM medium (Thermo Fisher™), was added into the remaining HEK293 cells to deactivate the Trypsin-EDTA. The harvested cells were stored in a -20°C freezer for subsequent DNA extraction. These were considered as the untreated or day 0-time point for analysis.

For depletion of mtDNA, cells were treated with EtBr (Inqaba Biotec) at day 3 and day 6. Once the cell growth confluency of 30 to 70 % was reached, EtBr treatment was performed. The HyClone™ DMEM media was supplemented with 0.25 µl of 10 mg/ml EtBr (Inqaba Biotec) to a final 50 ng/ml EtBr treatment. HEK293 cells were harvested (1 ml into a sterile tube) on days 3 and 6 after treatment (**TABLE 2.2**) and stored in -20°C freezer for subsequent DNA extraction. Experiments were performed in duplicate.

Table 2.7. Cell culture, treatment and harvest schedule.

	DAY 0	DAY 3	DAY 6
Cells Harvested for DNA Extraction	Yes	Yes	Yes
Treated with (EtBr)	No	Yes	Yes

DNA from HEK293 cells was extracted using Invitrogen™ DNA isolation Kit from Life Technologies™ (Thermo Fisher™). This kit precipitates and dehydrates protein by salting out (Miller *et al.*, 1988). Protease (125 µl) in nuclear lysis buffer was added to the pellet, vortexed for 5 seconds and then incubated on a 65°C heat block for 10 minutes and vortexed every 2 to 3 minutes. Then 275 µl of protein clearing solution was added to the sample and vortexed for 5 seconds, and incubated on ice for 10 minutes. The sample was centrifuged at 12000xg for 5 minutes, and the supernatant was carefully transferred into a labelled sterile 2 ml tube. Ice cold 90% ethanol (500 µl) was added to the new tube containing the DNA and the DNA was precipitated by inverting the tube multiple times until a white translucent mass was visible. The DNA pellet was centrifuged at 12000xg for 5 minutes and the supernatant was removed by decanting. Finally, 1 ml 70% ethanol was added to the DNA pellet and centrifuged for 2 minutes, after which the ethanol was removed. The DNA was dissolved in 150 µl of nuclease free water.

This procedure extracts both nDNA and mtDNA. The purity of the DNA was tested using a NanoDrop® ND-1000 Spectrophotometer. The DNA was stored at -20°C. Prior to use for qPCR the DNA was diluted to 20 ng/µl working concentration.

qPCR assays were run on the extracted DNA using primers in **Table 2.1** and reaction mix and cycling conditions shown in **Tables 2.4 and 2.5**. For each well, 2.5 µl of 20 ng/µl DNA was used. PCR reactions for the *GAPDH* and *NADH* genes were performed using DNA samples from each time point (i.e. DNA from Day 0, 3 and 6) in duplicate. PCR efficiency values obtained during primer optimisation were used for fold change calculations.

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 manually as $E^{\Delta\Delta CT}$ (Livak and Schmittgen, 2001), where

$$\Delta\Delta CT = [CT\,(target, untreated) - CT(ref, untreated)) - (CT(target, treated) - CT(ref, treated)]$$

where:
CT(target, untreated) = CT value of gene of interest (NADH1) in untreated sample
CT(ref, untreated) = CT value of control gene (GAPDH) in untreated sample
CT(target, treated) = CT value of gene of interest (NADH1) in treated sample
CT(ref, treated) = CT value of control gene (GAPDH) in treated sample
E = PCR efficiency

These results were used to confirm that our qPCR assay could detect the depletion of mtDNA induced by EtBr treatment and therefore validate the mtDNA copy number assay.

2.2.3 mtDNA copy number in cohort samples

qPCR reactions for each gene were run in duplicate in separate wells. For the total 263 samples, a total of 14 plates were used (i.e. 7 plates for each gene). Within each plate four samples of serial dilution (standards) were added. These standards were used to generate standard curves in each plate. A positive (calibrator) control sample (Day 0 DNA from HEK293 cells) and negative control samples were also used to account for any possible contamination, these were also duplicated on each plate. **Table 2.8** below depicts the plate map utilised in this study.

Table 2.8. Template of the 96-well plate map used in this study.

Plate 1	1	2	3	4	5	6	7	8	9	10	11	12
A	STD1	STD2	STD3	STD4	1	2	3	4	5	6	7	8
B	STD1	STD2	STD3	STD4	1	2	3	4	5	6	7	8
C	9	10	11	12	13	14	15	16	17	18	19	20
D	9	10	11	12	13	14	15	16	17	18	19	20
E	21	22	23	24	25	26	27	28	29	30	31	32
F	21	22	23	24	25	26	27	28	29	30	31	32
G	33	34	35	36	37	38	39	40	41	42	TC	NTC
H	33	34	35	36	37	38	39	40	41	42	TC	NTC

STD: Standards, **1-42:** Sample ID's, **TC:** Template Control (calibrator), **NTC:** No Template Control

qPCR assays were run on the cohort DNA samples using primer set 1 of both *NADH1* and *GAPDH* genes, reaction mix and cycling conditions as shown in **Tables 2.2 and 2.4**. For each sample, Ct values were obtained for each of the genes and were averaged across replicate wells. For calculation of mtDNA copy number per sample, $E^{\Delta CT}$ (Rooney *et al.*, 2015) was used, since we were not performing expression levels analysis in the current study.

The following quality control (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 from 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 GAPDH were excluded from analysis.

2.3 SNP choice (Bioinformatics analysis)

A thorough literature search was performed to identify SNPs of interest in the candidate genes. SNPs in candidate genes were chosen and prioritised according to the following criteria:

- Firstly, SNPs in candidate genes were chosen based on their reported or predicted functional effect or effect on mtDNA copy number and or association with SN (from literature review)
- only SNPs that reached MAF≥5% in African population Luhya Webuye in Kenya (LWK) were used
- It was noted whether SNPs were tag SNPs in African population Luhya Webuye in Kenya
- If SNP assay could be designed successfully in Sequenom software.

Each of these processes is described in more detail below.

2.3.1 Literature review of candidate genes

We used NCBI Pubmed (https://www.ncbi.nlm.nih.gov/pubmed), OMIM (https://www.omim.org) and SNPnexus (http://www.snp-nexus.org) to find literature where SNPs of interest in these genes had been previously described. SNPs of interest were those with:

- Effects on candidate gene expression
- Effects of candidate gene function
- Association with SN or any other neuropathy
- Association with mtDNA copy number

2.3.2 Minor allele frequency (MAF)

In the absence of genotype information from South African populations, SNPs with MAF ≥ 5% were chosen in the Luhya Webuye population in Kenya (LWK) as a proxy population. The following bioinformatics tools were used to check SNP frequencies in LWK:

- (NCBI) Variation viewer (http://www.ncbi.nlm.nih.gov/variation/view/)
- dbSNP (http://www.ncbi.nlm.nih.gov/SNP/)
- 1000 Genomes (http://browser.1000genomes.org/index.html) Entire Genome Interface for Exploring SNPs (ENGINES) (http://spsmart.cesga.es/about.php?dataSet=engines).
- International Hap Map Project (http://www.hapmap.org)

2.3.3 Predicted functional effect

We used Ensembl (http://www.ensembl.org/index.html) to confirm functional effects of SNPs of interest, since its database is embedded with various bioinformatics tools that can perform variant effect prediction (VEP). VEP is a powerful toolset for the analysis, annotation, and prioritization of genomic variants in coding and non-coding regions. VEP tools included Sorting Intolerant from Tolerant (SIFT) and PolyPhen-2; these scores were used in Ensembl to measure the severity of the consequence of the changes to protein coding sequences. SIFT scores range from 0.0 (deleterious) to 1.0 (tolerated) and PolyPhen-2 scores range from 0.0 (benign/tolerated) to 1.0 (deleterious).

2.3.4 SNP assay in Sequenom

Not all SNPs can be genotyped and or multiplexed when the MassARRay platform is employed. There are some SNPs that are difficult to genotype (e.g. repetitive region around the SNP of interest, SNPs that are close together and pseudogenes). The SNPs of interest were uploaded into Assay Design Suite v3.1 which searched for compatible primers. For every SNP, a forward, reverse and extension primer was designed. The program identified optimal primers to avoid

multiplexed primer combinations that may inappropriately extend primer amplification products from different SNPs.

2.3.5 Tag SNPs

SNPs chosen from the literature were checked to confirm whether they were tag SNPs or not. A tag SNP is a representative of other linked SNPs in a region of genome with linkage disequilibrium (LD) and the represented group of SNPs create a haplotype. Tag SNPs were only chosen if they tagged more than two other SNPs and had a minimum LD of $r^2 \geq 0.8$ with tagged SNPs. Tag SNPs were generated with Ensembl and Hap Map SNP genotype data in the African population of Luhya Webuye in Kenya.

Due to budgetary constraints, full tag SNP panels which would completely cover the LD patterns of each gene of interest could not be chosen. However, given the choice of several SNPs in a gene of equivalent functional interest, tag SNPs were prioritised over non-tag SNPs.

2.4 SNP genotyping

A total of 28 SNPs were genotyped in the cohort of 263 Black South Africans receiving HIV treatment. Two PCR multiplexes were sufficient for genotyping of all 28 SNPs in this cohort. **Table 2.9** below indicates which SNPs were in which multiplex.

Table 2.9. SNPs & their corresponding plex number during genotyping on the MassARRay spectrophotometer.

Gene	SNP ID	Plex #
PINK	rs10799655	1
	rs622525	1
ZFN648	rs7554182	1
NRF2	rs1962142	1
NFE2L3	rs4722585	1
TOP1mt	rs7387720	1
	rs724037	1
	rs2450772	1
TFAM	rs11006126	1
	rs2306604	1
	rs1049432	1
	rs11006132	1
C10orf2	rs17113613	1
	rs3740486	1
	rs11190787	1
LITAF	rs13333308	1
POLG2	rs9905016	1
	rs2075551	1
PINK	rs650616	2
CCDC19	rs2501325	2
NRF2	rs35652124	2
	rs2886162	2
TOP1mt	rs2293925	2
	rs11544484	2
	rs7460159	2
	rs4493919	2
C10orf2	rs3740485	2
POLG	rs2856268	2

Candidate SNPs in the cohort were genotyped using the Sequenom MassARRay by Inqaba Biotech (Pretoria). The Sequenom MassARRay platform employs the iPLEX Gold biochemistry (as illustrated by **Figure 2.2**). For each SNP, the designed forward and reverse PCR primers were used to generate a short PCR amplicon containing the SNP of interest using the following PCR reaction mix and cycling conditions (**Tables 2.10** and **Table 2.11**):

Table 2.10. Reaction components for PCR of target regions containing SNPs of interest

Component	Amount per PCR reaction (µl)
Nano-pure Water	0.8
10× PCR Buffer	0.5
25 mM $MgCl_2$	0.4
25 mM dNTP mix	0.1
0.5 µM Primer Pool	1.0
Polymerase, 5 U/µl	0.2
DNA template (25 ng/µl)	2.0

Table 2.11. Thermal cycler conditions for amplification of target regions

Step	Temperature	Incubation Time	Number of Cycles
1	94°C	2 minutes	1
2	94°C	30 seconds	46
	56°C	30 seconds	
	72°C	60 seconds	
3	72°C	5 minutes	1
4	4°C	∞	1

The unincorporated dNTPs were dephosphorylated by shrimp alkaline phosphatase (SAP) at 37°C for 40 minutes, followed by a 5-minute incubation period at 85°C to inactivate the enzyme. The samples then underwent another round of PCR using single base extension primers and mass-modified ddNTPs. The following PCR reaction mix and cycling conditions (**Table 2.12** and **Table 2.13**) for extension amplification were applied. Upon addition of chain terminators, there should be allele-specific differences in mass between extended PCR products. This mass difference allows the data analysis software to differentiate between SNP alleles as shown by **Figure 2.2** (http://agenabio.com/genetics).

Table 2.12. Reaction components for single base extension reaction.

Component	Amount per PCR reaction (µl)
Nano-pure Water	0.6182
10× iPLEX Buffer Plus	0.2
iPLEX dNTP Termination mix, 45×	0.2
Probe Pool (11 µM each)	0.9409
iPLEX Polymerase, 220×	0.0409
DNA template (PCR+SAP)	7.0

Table 2.13. Thermal cycling conditions of MassARRay instrument using iPlex Gold reaction mix.

		94°C	
60 Cycles/ ---->	5 Cycles ---		30s
		95°C	5s
		52°C	3s
		80°C	3s
		72°C	3 min
		15°C	HOLD

These products were then applied to a microchip for Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectrometry whereby a laser is used to generate ionized products. The ions accelerate into a tube towards the detector. The time of flight of these ionized molecules differ depending on the mass of the molecule, which is then measured by the mass spectrometer. The alleles were differentiated by their mass variation. The MassARRay analyser can detect DNA masses from 4500 to 9000 Daltons (Da) and can also discriminate analytes separated by atomic mass of 16 Da which is the smallest detectable difference between mass-modified chain terminators. Genotype calls were made in real-time during MALDI-TOF analysis. The cohort samples of 263 were genotyped using a total of three 96-well plates for genotyping in two multiplexes. The generated data were viewed using the data control software package SpectroTYPER™. This software translates the mass of the observed primers into a genotype for each reaction that allows the evaluation and management of the results which can be exported in Excel format.

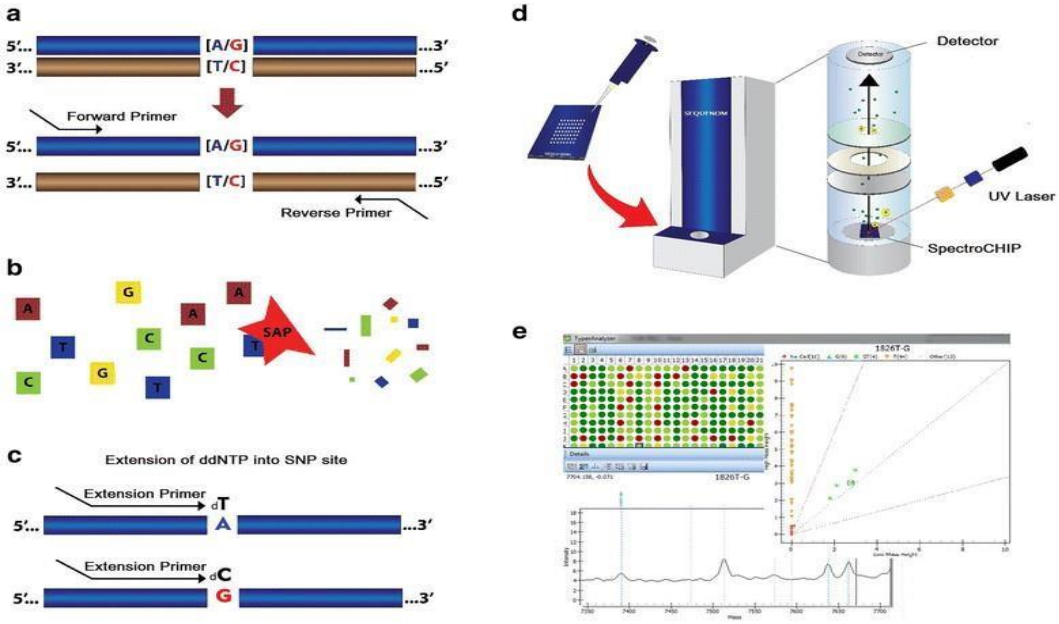


Figure 2.2. High throughput SNP genotyping work-flow with MassARRay spectrophotometer (Svidnicki et al., 2015). (a) PCR amplification of target of interest within the Genome, (b) Excess or unused primers are washed off from the amplicons, (C) Chain-terminators (ddNTPs) are added to the PCR reaction mix with amplicons, (d) The spectro-chip is loaded onto the MALDI-TOF for allele discrimination according to sizes generated by ddNTPs, (e) The TYPER Software is used for genotype calling (Homozygous or Heterozygous)

2.5 Statistical analysis

Plink v1.07 (Purcell *et al.*, 2007; http://zzz.bwh.harvard.edu/plink), HaploView v4.2 (Barrett *et al.*, 2005) and GraphPad Prism v7.0 (www.graphpad.com/scientific-software/prism/) statistical software were used for data analyses.

2.5.1 Analysis of association between demographic or clinical data, and SN

Descriptive statistical analyses were performed using GraphPad Prism v7.0. Demographic characteristics were described for continuous and categorical variables. Association between continuous data (such as age, months HIV-positive, months on d4T, CD4 count) and SN were analysed by t-test if parametric and by Mann-Whitney or Wilcoxon tests if non-parametric. Association between categorical data (such as gender) and SN were analysed by Chi-squared or Fishers exact tests P-value < 0.05 was considered significant. Any demographic or clinical phenotypes that were significantly associated with SN in this analysis, were used as covariates during SNP multivariate analysis as described below.

2.5.2 Analysis of association between demographic or clinical data, and mtDNA copy number

The associations between demographic variables and mtDNA copy number were performed using GraphPad Prism v7.0. In most cases both data sets consisted of continuous variables, therefore regression analyses were used. For gender, a categorical variable, the association with mtDNA copy number was examined using T-test.

2.5.3 Quality control (QC) of SNP data

Plink v1.07 (Purcell *et al.*, 2007) was used to analyse SNP data. Files of type *.ped, *.map and *..hlist were generated for use in Plink. Data quality was computed and the following standard parameters were used as QC limits as discussed in Laurie *et al.* (2010). Whatever data that failed to meet any of the settings listed below were excluded for analysis in this cohort.

Minor Allele Frequency (MAF)	<0.01
Genotyping Rate (< 85 %)	(--mind 0.15)
Hardy-Weinberg Equilibrium	< 0.001

2.5.4 Analysis of allele, genotype and haplotype frequencies in the whole cohort

After QC, allele and genotype frequencies in the entire cohort were calculated by direct observation. Where more than one SNP in a gene had been analysed, haplotypes and LD were calculated using Plink v1.07 and HaploView v4.20. Haplotypes of 2-3 SNPs were generated by Plink v1.07 using the sliding window approach. The LD blocks generated by HaploView used the plink.PED input file, and all LD blocks were generated with default settings in HaploView.

The allele frequencies in this South African cohort were compared to those in other African populations using data from 1000 Genomes via Ensembl (http://www.ensembl.org/Homo_sapiens/Info/Index). Data from a total of 120 individuals were used in each populations of LWK and Yoruba (YRI) (https://catalog.coriell.org/1/NHGRI/Collections/1000-GenomesCollections/1000-Genomes-Project). The allele frequencies of the SNPs of interest for the two African populations (LWK and YRI) were obtained using the 1000 genomes phase 3 data accessed via Ensembl database and HapMap release 28 accessed via Engines-SPSmart database (http://spsmart.cesga.es/about.php?dataSet=engines).

2.5.5 Analysis of SNP associations with phenotypes

Plink v1.07 was used to look for associations between candidate SNPs and clinical phenotypes.

Phenotypes that were considered included:

- SN (Where a patient presented with at least one symptom and one sign, as defined in Wadley *et al.,* 2013).
- Pain intensity (was measured on a scale of 0 to 10, with 0 being absence of pain and 10 representing extensive pain).
- mtDNA copy number (ratio of nDNA *GAPDH:* mtDNA *NADH1*)

Both univariate and multivariate analyses were conducted. The following model (**Figure 2.3**) was adopted from Hendry (2015), which follows Plink's standard

procedure for analysis of SNP data. All the selected SNPs and computed haplotypes were individually tested for association with each phenotypes using Plink v1.07. All associations that had $P < 0.05$ during univariate analysis were further tested for multivariate analysis which took into account any clinical or demographic phenotypes that were significantly associated with SN. In both univariate and multivariate analyses, $P < 0.05$ was considered to be significant.

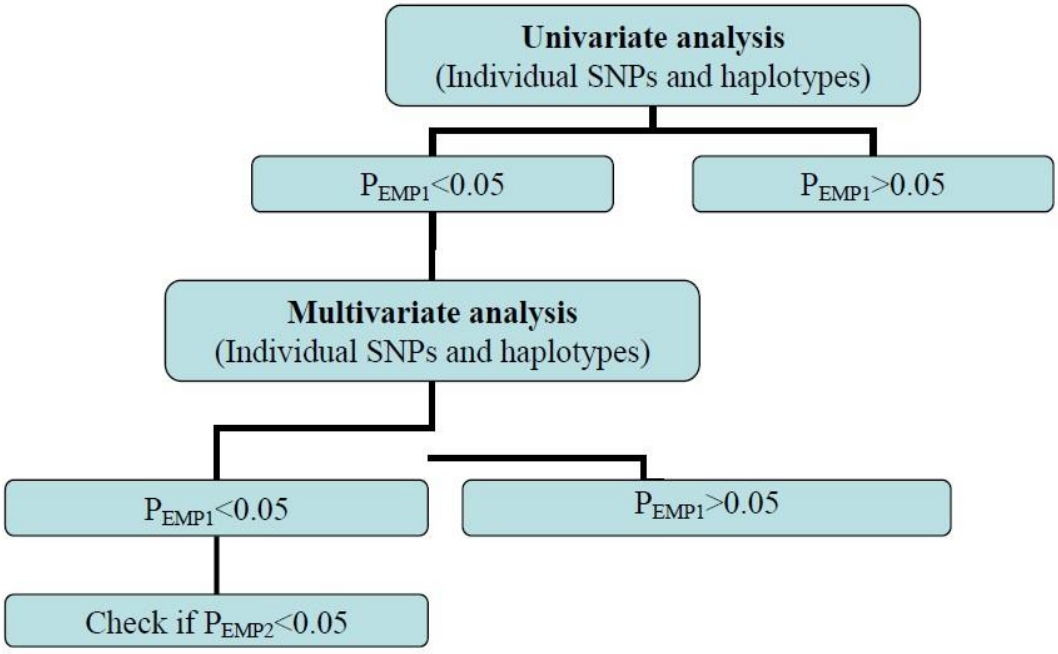


Figure 2.3: Model for analysis of data with the aid of Plink (Hendry, 2015).

2.5.5.1 Univariate analysis

The association of individual SNPs with categorical data such as SN presence or pain intensity were analysed using univariate data analysis methods of chi-square tests or Fisher's exact tests. Chi-square tests analysed associations in four models of inheritance namely, the allelic model and three genotypic models, including codominant, dominant and recessive models. Three different genotypic models of inheritance were used 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. If D is the minor allele and d is the major allele, the models were computed in the following manner (Purcell *et al.*, 2007):

Allelic	D	versus	d		
Dominant	(DD + Dd)	versus	dd		
Recessive	DD	versus	(Dd + dd)		
Genotypic (codominant)	DD	versus	Dd	versus	dd

Whenever five or less observations were presented in one of the models, Fisher's exact test was employed since chi-square test does not yield accurate results in this case. Fisher's exact test uses a 2 by 2 contingency table. Associations between haplotypes and phenotypes using Plink v1.07 (Sliding Window Approach) and HaploView v4.20 were also analysed. $P < 0.05$ was considered to be significant. Corrections for analysis of multiple SNPs were performed using the Bonferroni method (multiply P values by the number of tests performed or by the number of SNPs examined). This is used to reduce the chances of obtaining false-positive results (type I errors) when multiple pair wise tests are performed on a single set of data.

Additional empirical *P*-values (pEMP) were also generated using Plink v1.07, a Monte Carlo-based method for assessing significance of case–control association studies (http://pngu.mgh.harvard.edu/~purcell/plink/). One hundred thousand simulations were performed. Plink generated both uncorrected (EMP1) and corrected (EMP2) empirical P-values. The corrected empirical P-value (EMP2)

calculates the proportion of permutations in which any of the test statistics exceeded that particular observed statistic. It does not assume that the tests are independent, but 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).

2.5.4.2 Multivariate analysis

Only the SNPs that were found to be significant during univariate analyses were considered for multivariate analyses. Multivariate analysis was performed with Plink v1.07 using logistic regression for categorical outcomes or linear regression for continuous variables. Logistic regression analysis was employed for SN association with all SNPs of interest and haplotypes, since SN was a categorical variable. Pain intensity and mtDNA copy number association with all SNPs and haplotypes was calculated using linear regression analysis since pain intensity was a continuous variable.

The odds ratio (OR) and Beta values (standardized regression coefficients) presented in the results section were shown with respect to the minor allele in the sample for individual SNP association analysis. OR > 1 indicates that the minor allele increased the risk of development relative to the major allele. OR < 1 indicates a protective effect of the minor allele relative to the major allele. Beta values are a measure of how strongly each predictor variable influences the dependent variable. The beta is measured in units of standard deviation. For example, a beta value of 2.5 indicates that a change of one standard deviation in the predictor variable will result in a change of 2.5 standard deviations in the criterion variable. If a Beta coefficient is positive, then the relationship of this variable with the dependent variable is positive; this can also be interpreted as presence of minor allele association with an increased risk of developing a clinical phenotype. A negative Beta value indicates a negative relationship between dependent and independent variables, or a protective effect of the minor allele.

2.5.6 Genetic risk score (GRS)

A genetic risk score is an estimate of the cumulative contribution of genetic factors to a specific outcome of interest in an individual that takes into account a combination of reported risk alleles (Bailey and Igo, 2016). Although sometimes the GRS may also take into account the reported effect sizes. Effect sizes were not used in our calculations. While GRS is often based only on allele scores from significant loci, we found few significant loci in this study and therefore decided to attempt GRS using both significant ($P < 0.05$) and borderline significant ($P > 0.05$ but < 0.1) candidate loci P-values from univariate analysis. Each allele for each SNP was categorised as a risk allele or non-risk allele for the phenotype of interest, based on statistical findings and on literature review of SNP function (**Table 2.14**).

GRS scores were calculated manually for each individual. The score is simply a sum across SNPs of the number of risk alleles (0, 1 or 2) for each individual, therefore the minimum GRS score per individual was 0 and the maximum was (28 loci *2) = 56. We then analysed association between GRS score and phenotype using SNPs with $P < 0.1$. Chi-squared test, unpaired non-parametric T-test (Mann-Whitney test) and ordinary one-way ANOVA were used for association of SN, pain intensity and mtDNA copy number phenotypes respectively.

Table 2.14. Risk alleles and non-risk alleles of SNPs of interest in this study.

Gene	SNP	Allele 1 (RISK)	Allele 2
POLG	rs2856268	C	T
POLG2	rs9905016	T	C
	rs2075551	C	G
NRF2	rs35652124	C	T
	rs2886162	A	G
	rs1962142	A	G
C10orf2	rs17113613	A	G
	rs3740485	C	T
	rs3740486	T	C
	rs11190787	G	T
TOP1mt	rs2293925	A	G
	rs11544484	T	C
	rs2450772	A	G
	rs7387720	G	A
	rs724037	G	T
	rs7460159	G	A
	rs4493919	A	G
TFAM	rs2306604	A	G
	rs1049432	T	G
	rs11006132	G	A
	rs11006126	C	T
PINK	rs10799655	T	C
	rs650616	A	G
	rs622525	A	T
CCDC19	rs2501325	C	A
LITAF	rs13333308	C	T
ZFN648	rs7554182	T	C
NFE2L3	rs4722585	A	G

CHAPTER 3: RESULTS

3.1 Association of demographic characteristics with SN

Of the 263 participants, 54.5 % (n = 143) were clinically diagnosed with SN and the 120 individuals not clinically diagnosed with SN served as controls (45.5 %) Of those 143 individuals diagnosed with SN, 48.3% (n =127) reported pain as a symptom. Descriptive statistics for the cohorts used in this study are provided in **Table 3.1**. The demographic and clinical data collected from the patients were tested for significance between cases and controls. Age and height were significantly different in the cases vs. controls.

Table 3.1. Demographic characteristics of study participants.

	SN^+ n = 143	**SN^-** n = 120	**P-value**	**Total Cohort** n = 263
Age Mean(SD)	40.5(7.7)	35.86(7.2)	0.0001***	38.4
Height (cm) Mean(SD)	159.4(8.7)	156.4(7.4)	0.007***	158.05
CD4 Nadir (cells.mm^3) Mean (SD)	101.6(68.19)	100(74.63)	0.772	78.8
Gender (%)	**F**=77% **M**=23%	**F**=79% **M**=21%	0.766	**F**=78% **M**=22%
CD4 at Sampling Mean cells/mm^3 (SD)	440.7(199.6)	436.2(231.7)	0.446	426
Months HIV-positive Mean (SD)	55.39(41.15)	51.46(36.8)	0.550	47.8
Months on d4T Mean (SD)	17.34(25.13)	16.46(26.19)	0.934	16.9

SD : Standard Deviation
N : Total Samples
F : Female
M : Male
*** : Significant P-value

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

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

3.2 Cohort DNA quantification

DNA had already been extracted and stored at -20°C as several studies had already been conducted using these samples. For this study, the DNA was quantified for all remaining samples using the NanoDrop ND1000 spectrophotometer. The DNA from 263 participants was still acceptable for use with average DNA concentration of 594.95 ng/µl and average A260/280 value 1.85 (1.7-2.0). DNA samples were diluted to 20ng/µl aliquots.

3.3 qPCR assay

3.3.1 Conventional PCR and primer optimisation

PCR reactions for the nuclear reference gene (*GAPDH*) and mitochondrial target gene (*NADH1*) were optimised by trying two sets of primers for each gene. **Figure 3.1** depicts the results for *GAPDH* primer pair 2, which was found to be non-specific as shown by the double bands and the lower bands were less than expected size of 208 bp. The minimum annealing (Ta) temperature was 47°C and the maximum was 58°C. Primer pair 1 of *GAPDH* was found to be highly specific with a band of the correct size of 196 bp for all the annealing temperatures generated by the gradient PCR instrument (**Figure 3.2**). The Ta ranges were (45-50) °C. The final Ta chosen was 50°C (**Figure 3.2**), this temperature was then used for the qPCR.

The two *NADH1* primer pairs yielded single bands of the correct sizes of 154 bps and 243 bps respectively during optimisation (**Figure 3.3**). The negative control was also clear of any contaminants and primer-dimers were visible on the gel image. The final primer volume was decreased to 0.6 µl at 10uM concentration since it produced no primer dimers. Primer pair 1 (Ta = 58°C) was chosen for further optimisation on the qPCR since it had a smaller amplicon size of 154 bps. This also was in correspondence with the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) which suggests small amplicons (Bustin *et al.,* 2009).

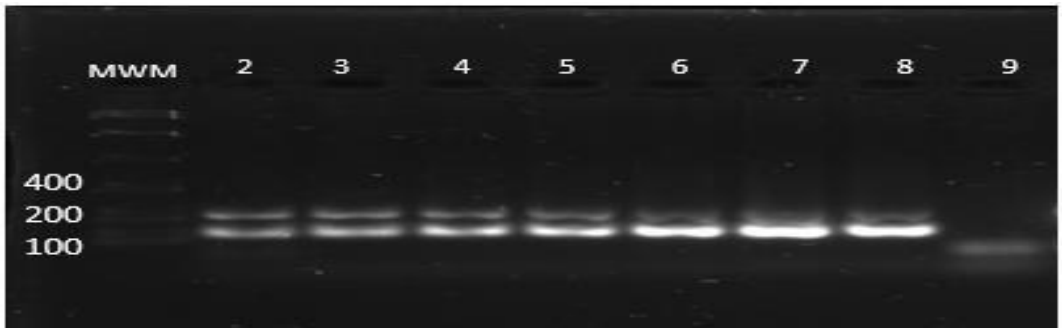


Figure 3.1: Optimisation results of *GAPDH* primer set 2. Lane 1 (Molecular weight marker), Lane 2-8 (DNA samples) and Lane 9 (Non-template control). The minimum annealing (Ta) temperature was 47°C and the maximum was 58°C.

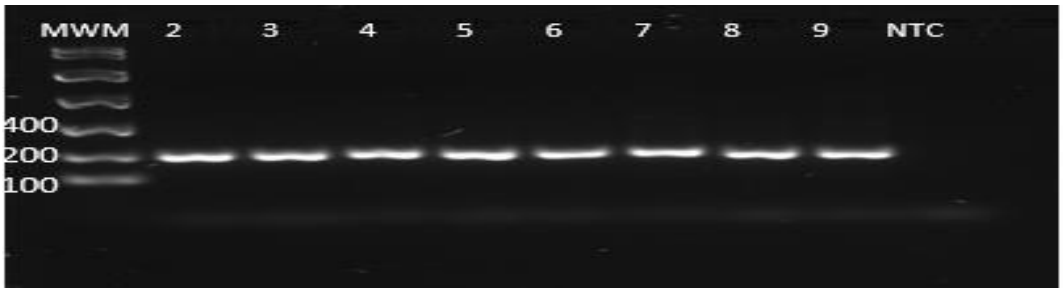


Figure 3.2: optimisation results of *GAPDH* primer 1. The Ta ranges were (45-50) °C. The final Ta chosen was 50°C (Lane 2), this temperature was then used for the qPCR. Lane 1 (Molecular weight marker), Lane 2-9 (DNA samples) and Lane 10 (Non-template control).

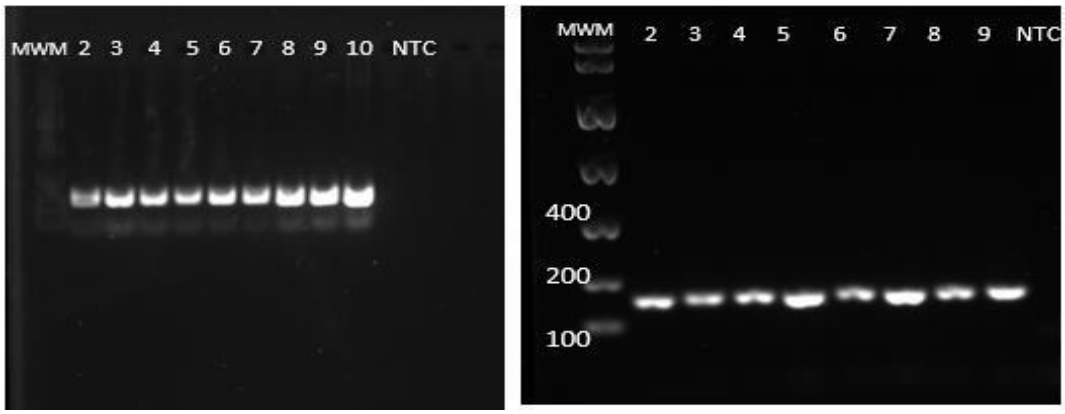


Figure 3.3: Optimisation results of *NADH1*. (A) Primer pair 2, (B) Primer pair 1. Lane 1 (Molecular weight marker), Lane 2-9 (DNA samples) and Lane 10 (Non-template control). Ta ranges pair 1 (48-59) °C and pair 2 (45-50) °C.

3.3.2 qPCR optimisation

The PCR primers were optimised for qPCR using SYBR Green chemistry for detection. Standard and melting curves were generated. **Figure 3.4** depicts the results obtained. The standard curves were generated so that we could calculate how efficient the PCR reactions were when amplifying the target region. Efficiency rate (E%) between 90 and 110% (Bio-rad) or 1.90 to 2.10 values (Roche) was considered acceptable (Taylor *et al.,* 2010; Pettengil *et al.,* 2012). Both the target and reference genes efficiency (E) was within the recommended range as efficiencies of 2.04 and 1.93 were obtained for *NADH1* and *GAPDH* primers respectively (in Roche terminology). The efficiencies were both close enough to consider the usage of the Livak equation (Livak and Schmittgen, 2001). The melting curve analysis further supported the specificity seen with normal PCR in **Figures 3.2 and 3.3**. Only one peak was visible on the melting curves of both primer sets thus proving the specificity.

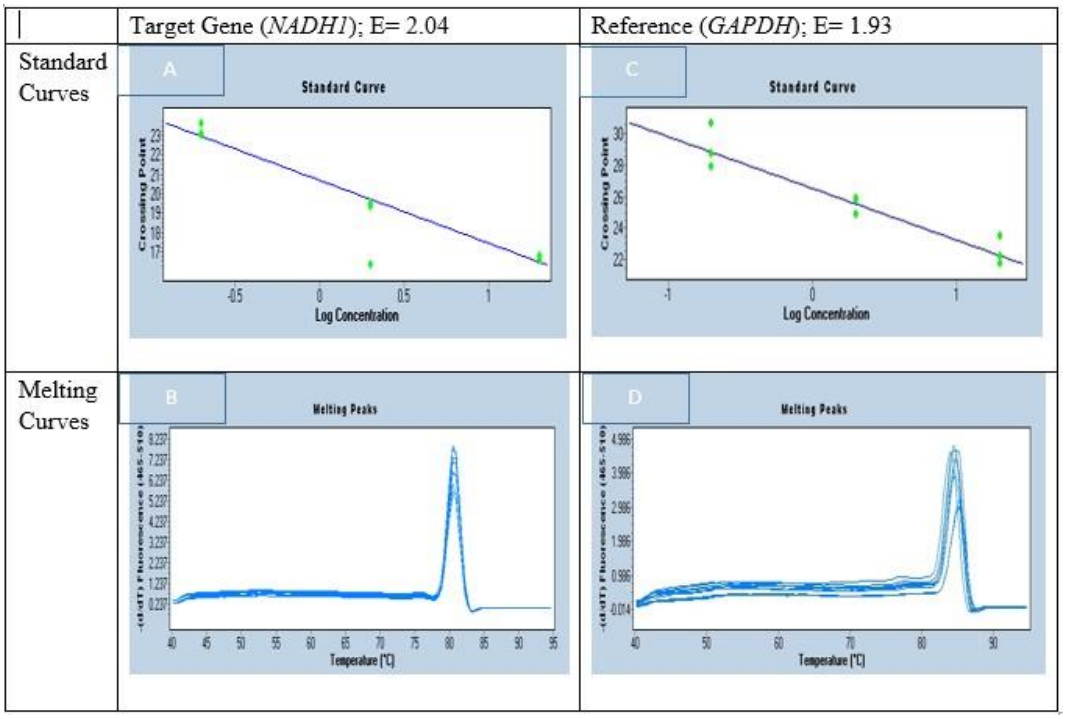


Figure 3.4. qPCR primer optimisation of both target and reference genes.
(A+C) = standard curves with 2 replicates per sample. (B+D) = melting curves with melting temperature (Tm) of 55 °C for *GAPDH* and 63 °C for *NADH1*.

3.3.3 qPCR assay validation

Table 3.2 shows the threshold cycle (Ct) values obtained for the target and reference genes in the validation samples treated with EtBr. The obtained Ct values were computed using the Light Cycler 480 SW 1.5 software. Averages of duplicated samples (replicates) were calculated (**Table 3.3**). The Ct values of the target mtDNA gene *NADH1* increased from day 0 to day 3, and from day 3 to day 6, indicating decreasing mtDNA, and the Ct values of the reference nuclear gene *GAPDH* remained fairly constant for all time points. The comparison between the target gene Ct values and reference gene Ct values is depicted by **Figure 3.5**.

Table 3.3. qPCR threshold cycle (Ct) values obtained using HEK293 cells DNA.

Day	Target (*NADH1*)			Reference (*GAPDH*)		
	Replicate Ct Values		Average Ct	Replicate Ct Values		Average Ct
0 (Untreated)	14.28	16.00	15.14	23.08	23.51	23.3
3 (Treated with EtBr)	16.82	17.52	17.17	22.89	22.58	22.7
6 (Treated with EtBr)	19.17	19.87	19.52	22.91	22.87	22.9

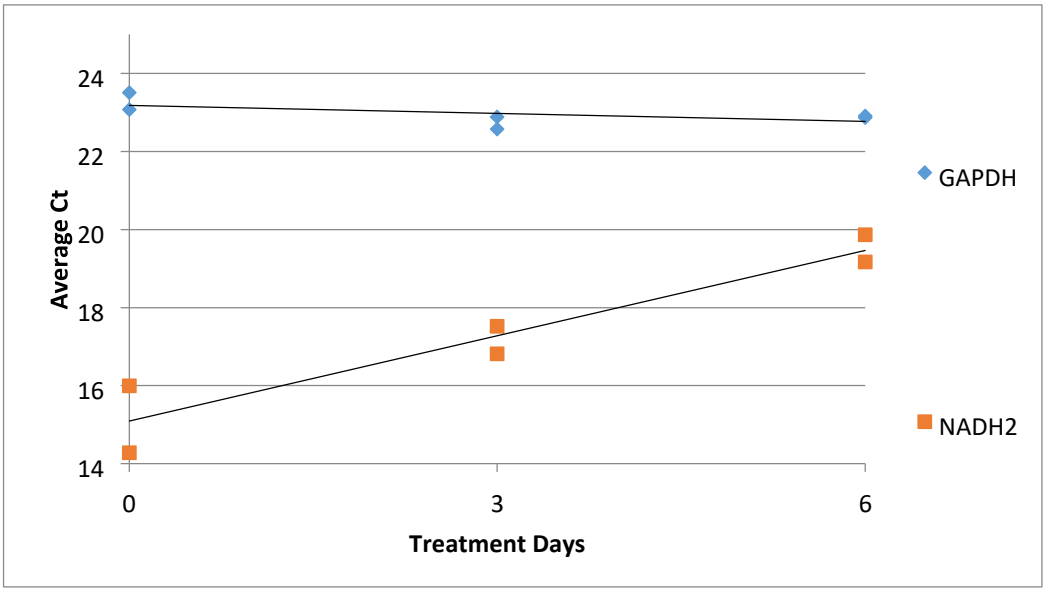


Figure 3.5. Comparisons of average Ct values of *NADH1* & *GAPDH* genes over 6 days.

Table 3.4 shows the fold changes of *NADH1* from day 0 to day 6, relative to *GAPDH1*. Day 0 served as a reference since the cells on that day they were not treated with EtBr. The Livak equation used for calculation of fold change of nDNA:mtDNA ratio between EtBr treated and untreated was $E^{\Delta\Delta CT}$.

The mtDNA: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 *NADH1* gene compared to day0, while on day 6 the fold change increased to 27.67-fold compared to day 0. On Day 6 we also observed extensive cell death (the medium in which the cells were cultured was turning cloudy). 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 this.

Table 3.4. Fold change ratios of *NADH1* relative to *GAPDH*.

Days	**ΔΔCT** (*GAPDH –NADH*)	**$E^{\Delta\Delta CT}$ (fold change)** (Treated compared to untreated)
(0-3)	2.59	6.02
(3-6)	2.195	4.57
(0-6)	4.79	27.67

3.3.4 qPCR in cohort samples

Once QC settings to the qPCR results was complete, results from 206 individuals remained for further analysis in both target and reference genes (**Table 3.5**) including cases (n = 135) and controls (n = 71).

Our review of the qPCR results showed that negative controls on each plate had CT values of 35 or higher. In most cases, CT values for the *NADH1* gene were higher than CT values for the *GAPDH* gene, suggesting that as expected the amount of mtDNA exceeded the amount of nDNA. **Table 3.5** shows the mean CT, mean ΔCT and mean $E^{\Delta CT}$ (mtDNA:nDNA ratio) values for the cases and controls respectively. No significant difference in $E^{\Delta Ct}$ values (**Table 3.5**) was observed between cases and controls (P >0.05). **Figure 3.6** shows mtDNA copy number distribution in both cases and controls and no significance was detected with P = 0.24.

The Livak method was employed for the calculation of the fold change ratio of our target gene (*NADH1*) in our target (cases) sample relative to our control cohorts by calculating $2^{\Delta\Delta CT}$. The fold change was not extensive since there was only 1.21-fold decrease in mtDNA:nDNA ratio in patients with SN as compared to those without SN.

Table 3.5. Average $2^{\Delta Ct}$ fold difference in cases vs controls and exact P-value.

	SN+ (n = 135)	**SN$^-$** (n = 71)	**P-value**
Mean CT for GAPDH	21.19	21.21	-
Mean CT for NADH	13.99	13.74	-
Mean ΔCT	7.20	7.47	
Mean $2^{\Delta CT}$ (SD) mtDNA: nDNA ratio	197(159.1)	223.7(156.3)	0.099
ΔΔCT	0.27		-
Fold change = $2^{\Delta\Delta CT}$	1.21		-

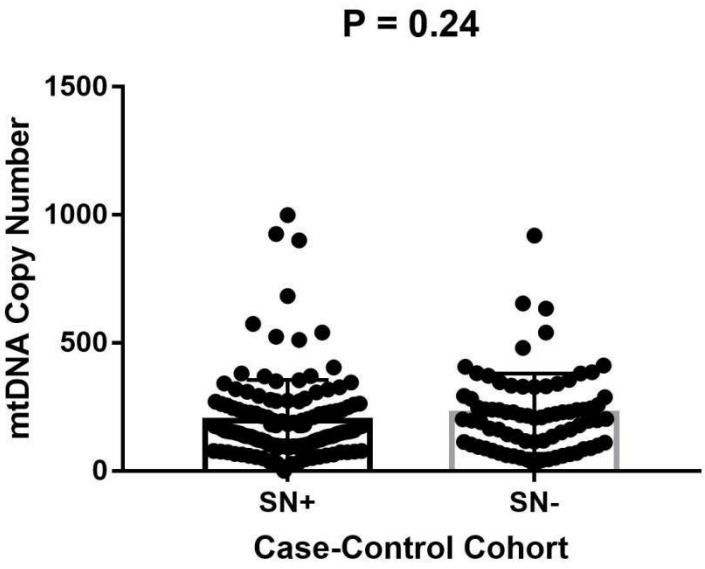


Figure 3.6. mtDNA copy number (*NADH1* to *GAPDH*) was not significantly different between SN+ cases vs. SN- controls.

3.4 SNP selection in genes of interest.

SNPs in candidate genes were chosen and prioritised according to the criteria described in Methods and Materials.

We identified a total of 28 SNPs of interest to be genotyped (**Table 3.6**), of which 24 were from the mtDNA replication pathway, and four were from Leger et al. (2014). A total of 24 SNPs had associated functional effects, and four SNPs (*POLG2* rs2075551, *C10orf2* rs11190787, TOP*1mt* rs7460159 and rs4493919) were chosen purely on the basis of being promoter region tag SNPs (no function known).

The following SNP (*NRF2* rs6721961) was excluded due to inability to be included in the multiplex designed with other SNPs.

Table 3.6. List of SNPs chosen for genotyping in this cohort.

GENE	SNP ID (Alternative ID)	Alleles	Minor Allele (Risk)	MAF LWK	Tag SNP (number of other SNPs tagged)	Variant Type	Functional Effect	SIFT / PolyPhen-2 score	References
POLG	rs2856268 T (-365) C	T/C	C	0.19	Yes (4)	Regulatory region / promoter variant 5'UTR	Changes recognition site for AP1 transcription factor; Low POLG Expression; Higher POLG mRNA in breast cancer; Diabetic Polyneuropathy.	-	Popanda et al. 2013 Spitsina *et al.*, 2009 Malyarchuk *et al.*, 2011
POLG2	rs9905016	C/T	T	0.26	Yes (1)	5'UTR/Upstream gene variant	Increased Transcription of TC compared to TT Increased risk of oral cancer TT associates with lower mtDNA levels in oral cancer than CT or CC	-	Datta *et al.*, 2016
	rs2075551	C/G	C	0.28	Yes (1)	5'UTR variant	Unknown	-	ENSEMBL
NRF2	rs35652124	T/C	C	0.11	Yes (4)	Intron Promoter	Reduced binding affinity of transcription factor; Decreased Parkinson disease	-	Von Otter *et al.*, 2014
	rs2886162	G/A	A	0.46	Yes (2)	Intron	Low NRF2 mRNA eExpression	-	Hartikainen *et al.*, 2012
	rs1962142	G/A	A	0.06	No	Intron	Low NRF2 protein cytoplasmic eExpression; Associated with neuropathy in diabetes	-	Hartikainen *et al.*, 2012 Xu 2016

GENE	SNP ID (Alternative ID)	Alleles	Minor Allele (Risk)	MAF LWK	Tag SNP (number of other SNPs tagged)	Variant Type	Functional Effect	SIFT / PolyPhen-2 score	References
C10orf2	rs17113613 (c.1102G>A)	G/A	A	0.165	No	Missense (V368, p. V368Ile) Exon 1	Affects subunit interactions; decreases mtDNA, Associated with muscle weakness	0.45 (Tolerated) SIFT	Singh *et al.*, 2012; Naimi et al. 2006
	rs3740485	C/T	C	0.45	Yes (4)	Splice region variant	Progressive External Opthalmoplegia (mtDNA copy number)	-	Singh *et al.*, 2012
	rs3740486	T/C	T	0.45	Yes (4)	Splice region variant	Progressive External Opthalmoplegia (mtDNA copy number)	-	Singh *et al.*, 2012
	rs11190787	G/T	G	0.45	Yes (4)	5'UTR variant	Unknown	-	ENSEMBL
TOP1mt	rs2293925 (g.144392368 G>A)	G/A	A	0.08	Yes	Missense Arg525Trp	Mitochondrial disorders (PEO and Alpers syndrome)	0 (Damaging) SIFT	Wang *et al.*, 2011
	rs11544484 (g.144406705 C>T)	C/T	T	0.49	Yes (4)	Missense V256I Exon 6	Mitochondrial disorders (PEO and Alpers syndrome)	1 (Tolerated) SIFT	Wang *et al.*, 2011
	rs2450772 (g.144413416 G>A)	G/A	A	0.33	Yes (4)	Missense D216D Exon 2	Mitochondrial disorders (PEO and Alpers syndrome)	-	Wang *et al.*, 2011
	rs7387720	A/G	G	0.26	Yes (4)	Downstream Gene Variant	Mitochondrial dysfunction in patients with Diabetic kidney disease.	-	Swan *et al.*, 2015

GENE	SNP ID (Alternative ID)	Alleles	Minor Allele (Risk)	MAF LWK	Tag SNP (number of other SNPs tagged)	Variant Type	Functional Effect	SIFT / PolyPhen-2 score	References
TOP1mt	rs724037	T/G	G	0.30	Yes (4)	Intron Variant	Mitochondrial dysfunction in patients with Diabetic kidney disease.	-	Swan *et al.*, 2015
	rs7460159	A/G	G	0.09	Yes	5 prime UTR variant	Unknown	-	ENSEMBL
	rs4493919	G/A	A	0.24	Yes	5 prime UTR variant	Unknown	-	ENSEMBL
TFAM	rs2306604	G/A	A	0.15	Yes (2)	Intron Variant	Cryptic Splice Site = Truncated Protein. mtDNA Depletion in Neurodegenerative Disorders. Association with Alzheimer's Disease and Parkinson's disease	-	Belin *et al.*, 2007 Gatt *et al.*, 2013 Migliore & Coppedè, 2009
	rs1049432	G/T	T	0.36	Yes (4)	3'UTR variant	Mitochondrial dysfunction in patients with HD. Associated with Huntington's Disease risk	-	Taherzadeh-Fard *et al.*, 2011
	rs11006132	A/G	G	0.37	Yes (4)	3'UTR variant	Associated with Huntington's Disease risk	-	Taherzadeh-Fard *et al.*, 2011
	rs11006126	T/C	C	0.41	No	3'UTR variant	Low mtDNA Copy Number in Blood cells.	-	Cai *et al.*, 2015

GENE	SNP ID (Alternative ID)	Alleles	Minor Allele (Risk)	MAF LWK	Tag SNP (number of other SNPs tagged)	Variant Type	Functional Effect	SIFT / PolyPhen-2 score	References
PINK1	rs10799655	C/T	T	0.33	No	Upstream gene variant	Increased PINK1 mRNA	-	Franks *et al.*, 2008
	rs650616	G/A	A	0.44	No	Non-coding transcript exon variant	Increased PINK1 mRNA	-	Franks *et al.*, 2008
	rs622525	T/A	A	0.14	No	Intron	Increased PINK1 mRNA	-	Franks *et al.*, 2008
CCDC19	rs2501325	A/C	C	0.23	Yes (3)	Intron Variant	Sensory Neuropathy	-	Leger *et al.*, 2014
LITAF	rs13333308	T/C	C	0.11	Yes (3)	Intron Variant	Sensory Neuropathy	-	Leger *et al.*, 2014
ZNF648	rs7554182	C/T	T	0.27	No	Downstream gene variant	Sensory Neuropathy	-	Leger *et al.*, 2014
NFE2L3	rs4722585	G/A	A	0.27	No	Regulatory region variant	Sensory Neuropathy	-	Leger *et al.*, 2014

3.5 Sequenom genotyping results

Figure 3.7A shows an example of a MassARRay spectrum multiplexing the candidate SNPs chosen for genotyping in this study. From this MassARRay Spectrum, using the TYPER Software a report was generated which distinguished between SNP alleles (homozygous or heterozygous) within each sample interrogated. Here SNP rs9905016 from POLG 2 was used as an example.

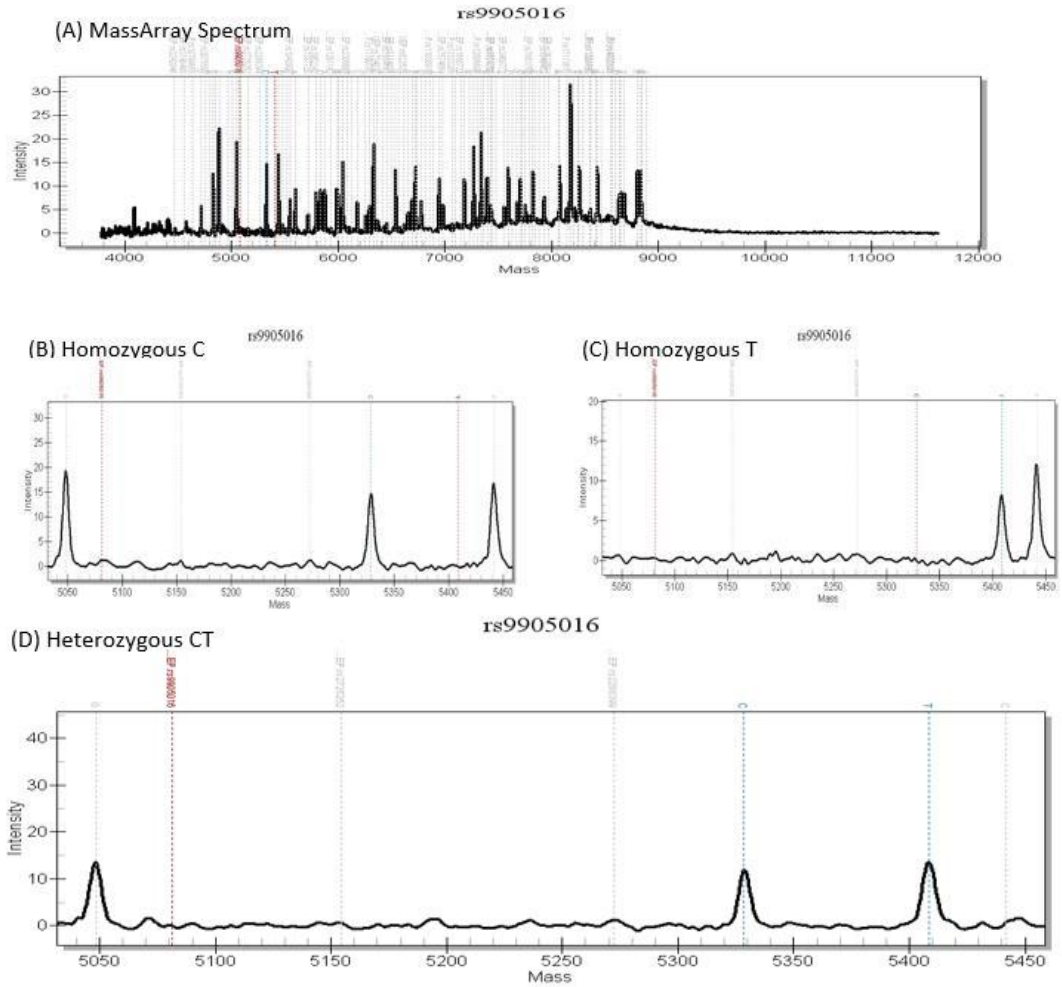


Figure 3.7. High throughput genotyping data from the MassARRay system showing results from multiplexed SNPs distributed across the range of detectable DNA masses from 4500 to 9000 Da.
SNP rs9905016 genotypes highlighted in red on the MassARRay spectrum.
(A) Homozygous for the CC allele with mass of 5330 Da only,
(B) Homozygous for the T allele with mass of 5410 Da only
(C) Heterozygous for the CT allele with mass of both 5330 and 5410 Da.

The genotyping data from n = 263 and 28 SNPs generated by Sequenom MassARRay was filtered using the Plink quality control settings listed in the Methods section. The total number of samples was reduced from 263 to 248, due to the removal of individuals who exhibited <85 % genotyping rate. Of the total 28 SNPs, 26 SNPs remained after QC (**Table 3.7**). The two missing SNPs (*TOP1mt* rs7460159 and rs4493919) failed the Hardy-Weinberg Equilibrium and data missingness test, so they were filtered out and were not included.

Table 3.7. Plink QC data of whole cohort samples.

	Quality Control (QC)	
	Unfiltered Data	**Filtered Data**
Total Cohort Samples	263	248
Cases	143	136
Controls	120	112
Males	58	54
Females	205	194
Total SNPs	28	26

3.6 Allele frequencies in whole cohort vs. other African populations

Allele frequencies were calculated in our total cohort (**Table 3.8**). 25 SNPs had MAF ≥ 0.05, with MAF ranging from 0.064 to 0.438 in SA cohort. *TOP1mt* SNP rs2293925 had a very low MAF of 0.012. This is the first report of these SNP MAF in a South African cohort.

APPENDIX C shows the genotype frequencies of all 26 SNPs that passed QC settings in the cohort of 248 individuals. In all the controls, Hardy-Weinberg Equilibrium (HWE) P-values were above 0.05.

We then compared the SA MAF to those in LWK and YRI populations (from 1000 genomes data), two African populations that may share common ancestry with South African populations. The frequencies vary across all three populations regardless of all being located on the African continent.

Table 3.8. Distribution of allele frequencies in three African population groups, [Kenya (LWK) and Nigeria (YRI)] and South Africans.

GENE	SNP ID	Minor Allele	LWK n = 120	YRI n = 120	SA (this study) n = 263	P-value
			MAF	MAF	MAF	
POLG	rs2856268	C	0.17	0.17	0.14	0.799
POLG2	rs9905016	T	0.26	0.19	0.17	0.256
	rs2075551	C	0.28	0.25	0.20	0.412
NRF2	rs35652124	C	0.11	0.21	0.18	0.150
	rs2886162	A	0.46	NG	0.44	NA
	rs1962142	A	0.06	0.06	0.06	0.999
C10orf2	rs17113613	A	0.17	0.10	0.14	0.352
	rs3740485	C	0.45	0.46	0.38	0.461
	rs3740486	T	0.45	0.46	0.38	0.461
	rs11190787	G	0.45	0.46	0.37	0.370
TOP1mt	rs2293925	A	0.08	0.01	0.01	0.006
	rs11544484	C	0.49	0.43	0.43	0.616
	rs2450772	A	0.33	0.43	0.39	0.342
	rs7387720	G	0.26	0.21	0.21	0.622
	rs724037	G	0.30	0.23	0.20	0.239
	rs7460159	G	0.09	0.02	0.38	0.0001
	rs4493919	A	0.24	0.19	0.19	0.602
TFAM	rs2306604	A	0.15	0.15	0.21	0.427
	rs1049432	T	0.36	0.32	0.22	0.084
	rs11006132	G	0.37	0.33	0.23	0.089
	rs11006126	C	0.41	0.36	0.31	0.338
PINK	rs10799655	T	0.33	0.38	0.30	0.481
	rs650616	A	0.44	0.43	0.42	0.960
	rs622525	A	0.14	0.27	0.18	0.060
CCDC19	rs2501325	C	0.23	0.14	0.14	0.148
LITAF	rs13333308	C	0.11	0.13	0.10	0.793
ZNF648	rs7554182	T	0.27	0.24	0.26	0.885
NFE2L3	rs4722585	A	0.27	0.24	0.26	0.885

*NG : Not Genotyped.

3.7 Haplotypes and linkage disequilibrium (LD) in cohort

A total of 26 SNPs in eleven genes were examined in this study. For six of these genes, more than one SNP per gene was studied and therefore haplotypes in the following genes could be considered: *TOP1mt* (3 SNPs), *C10orf2* (4 SNPs), *TFAM* (4 SNPs), *PINK* (3 SNPs), *NRF2* (3 SNPs) and *POLG2* (2 SNPs). In addition, some of the genes were on the same chromosome, notably *PINK*, *ZNF648* and *CCDC19* were all on chromosome 1, and *TFAM* and *C10orf2* were both on chromosome 10, so haplotypes across these genes could also be examined.

3.7.1 Haplotype frequencies in cohort

Haplotype frequencies were generated with Plink v1.07 software using the sliding window approach. All possible three SNP haplotypes on each chromosome were generated using the 26 SNPs dataset. A total of 11 possible 3-locus combinations (windows) were computed along with their frequencies in the cohort samples (**Table 3.9**). There were eleven 3-locus combinations (windows) that were considered but many more possible haplotypes than 11 because each 3-locus combination considered can have several different haplotypes. So, a total of 62 possible haplotypes were generated.

Table 3.9. Frequencies of 3 SNP haplotypes in whole cohort.

LOCUS	CHROMOSOME	GENE	SNPs	HAPLOTYPE	FREQUENCY
WIN1	1	*PINK*	rs10799655|rs650616|rs622525	TGA	0.173
WIN1	1	*PINK*	rs10799655|rs650616|rs622525	CAT	0.422
WIN1	1	*PINK*	rs10799655|rs650616|rs622525	TGT	0.124
WIN1	1	*PINK*	rs10799655|rs650616|rs622525	CGT	0.277
WIN2	1	*PINK – CCDC19*	rs650616|rs622525|rs2501325	GAC	0.010
WIN2	1	*PINK – CCDC19*	rs650616|rs622525|rs2501325	ATC	0.052
WIN2	1	*PINK – CCDC19*	rs650616|rs622525|rs2501325	GTC	0.073
WIN2	1	*PINK – CCDC19*	rs650616|rs622525|rs2501325	GAA	0.167
WIN2	1	*PINK – CCDC19*	rs650616|rs622525|rs2501325	ATA	0.370
WIN2	1	*PINK – CCDC19*	rs650616|rs622525|rs2501325	GTA	0.328
WIN3	1	*PINK – CCDC19-ZFN*	rs622525|rs2501325|rs7554182	TCT	0.025
WIN3	1	*PINK – CCDC19-ZFN*	rs622525|rs2501325|rs7554182	AAT	0.040
WIN3	1	*PINK – CCDC19-ZFN*	rs622525|rs2501325|rs7554182	TAT	0.180
WIN3	1	*PINK – CCDC19-ZFN*	rs622525|rs2501325|rs7554182	TCC	0.101
WIN3	1	*PINK – CCDC19-ZFN*	rs622525|rs2501325|rs7554182	AAC	0.129
WIN3	1	*PINK – CCDC19-ZFN*	rs622525|rs2501325|rs7554182	TAC	0.517
WIN4	2	*NRF2*	rs1962142|rs35652124|rs2886162	ATA	0.053
WIN4	2	*NRF2*	rs1962142|rs35652124|rs2886162	GTA	0.384
WIN4	2	*NRF2*	rs1962142|rs35652124|rs2886162	GCG	0.178
WIN4	2	*NRF2*	rs1962142|rs35652124|rs2886162	GTG	0.373
WIN6	8	*TOP1mt*	rs7387720|rs2293925|rs724037	GAG	0.011
WIN6	8	*TOP1mt*	rs7387720|rs2293925|rs724037	GGG	0.181
WIN6	8	*TOP1mt*	rs7387720|rs2293925|rs724037	AGG	0.017
WIN6	8	*TOP1mt*	rs7387720|rs2293925|rs724037	GGT	0.019
WIN6	8	*TOP1mt*	rs7387720|rs2293925|rs724037	AGT	0.772
WIN7	8	*TOP1mt*	rs2293925|rs724037|rs11544484	AGC	0.012
WIN7	8	*TOP1mt*	rs2293925|rs724037|rs11544484	GGC	0.191

LOCUS	CHROMOSOME	GENE	SNPs	HAPLOTYPE	FREQUENCY
WIN7	8	*TOP1mt*	rs2293925|rs724037|rs11544484	GTC	0.225
WIN7	8	*TOP1mt*	rs2293925|rs724037|rs11544484	GTT	0.570
WIN8	8	*TOP1mt*	rs724037|rs11544484|rs2450772	GCA	0.139
WIN8	8	*TOP1mt*	rs724037|rs11544484|rs2450772	TCA	0.015
WIN8	8	*TOP1mt*	rs724037|rs11544484|rs2450772	TTA	0.228
WIN8	8	*TOP1mt*	rs724037|rs11544484|rs2450772	GCG	0.065
WIN8	8	*TOP1mt*	rs724037|rs11544484|rs2450772	TCG	0.208
WIN8	8	*TOP1mt*	rs724037|rs11544484|rs2450772	TTG	0.342
WIN9	10	*TFAM*	rs11006126|rs2306604|rs1049432	CGT	0.213
WIN9	10	*TFAM*	rs11006126|rs2306604|rs1049432	TGT	0.011
WIN9	10	*TFAM*	rs11006126|rs2306604|rs1049432	CAG	0.053
WIN9	10	*TFAM*	rs11006126|rs2306604|rs1049432	TAG	0.150
WIN9	10	*TFAM*	rs11006126|rs2306604|rs1049432	CGG	0.047
WIN9	10	*TFAM*	rs11006126|rs2306604|rs1049432	TGG	0.527
WIN10	10	*TFAM*	rs2306604|rs1049432|rs11006132	GTG	0.224
WIN10	10	*TFAM*	rs2306604|rs1049432|rs11006132	AGA	0.203
WIN10	10	*TFAM*	rs2306604|rs1049432|rs11006132	GGA	0.565
WIN11	10	*TFAM-C10orf2*	rs1049432|rs11006132|rs17113613	TGA	0.036
WIN11	10	*TFAM-C10orf2*	rs1049432|rs11006132|rs17113613	GAA	0.104
WIN11	10	*TFAM-C10orf2*	rs1049432|rs11006132|rs17113613	TGG	0.188
WIN11	10	*TFAM-C10orf2*	rs1049432|rs11006132|rs17113613	GAG	0.665
WIN12	10	*TFAM-C10orf2*	rs11006132|rs17113613|rs3740485	GAC	0.035
WIN12	10	*TFAM-C10orf2*	rs11006132|rs17113613|rs3740485	AAC	0.105
WIN12	10	*TFAM-C10orf2*	rs11006132|rs17113613|rs3740485	GGC	0.062
WIN12	10	*TFAM-C10orf2*	rs11006132|rs17113613|rs3740485	AGC	0.191
WIN12	10	*TFAM-C10orf2*	rs11006132|rs17113613|rs3740485	GGT	0.133
WIN12	10	*TFAM-C10orf2*	rs11006132|rs17113613|rs3740485	AGT	0.474
WIN13	10	*C10orf2*	rs17113613|rs3740485|rs3740486	ACT	0.141
WIN13	10	*C10orf2*	rs17113613|rs3740485|rs3740486	GCT	0.245
WIN13	10	*C10orf2*	rs17113613|rs3740485|rs3740486	GTC	0.609

LOCUS	CHROMOSOME	GENE	SNPs	HAPLOTYPE	FREQUENCY
WIN14	10	*C10orf2*	rs3740485|rs3740486|rs11190787	CTG	0.379
WIN14	10	*C10orf2*	rs3740485|rs3740486|rs11190787	TCT	0.609
WIN17	17	*POLG2*	rs9905016|rs2075551	TC	0.169
WIN17	17	*POLG2*	rs9905016|rs2075551	CC	0.033
WIN17	17	*POLG2*	rs9905016|rs2075551	CG	0.799

3.7.2 LD analysis

With the aid of HaploView v4.2 software, linkage disequilibrium was assessed within the six genes with > 1 SNP, *TOP1mt* (5 SNPs), *C10orf2* (4 SNPs), *TFAM* (4 SNPs), *PINK* (3 SNPs), *NRF2* (3 SNPs) and *POLG2* (2 SNPs). Haplotype blocks per gene were generated (**Figure 3.8**) and due to few markers being interrogated per gene, the haplotype blocks covered small regions of the entire gene of interest. Many but not all of the SNPs within each gene were in strong LD. All studied SNPs in *C10orf2*, *PINK1* and *POLG2* were in strong LD whereas different linkage blocks were observed for the SNPs studied in *TOP1mt*, *TFAM* and *NRF2*.

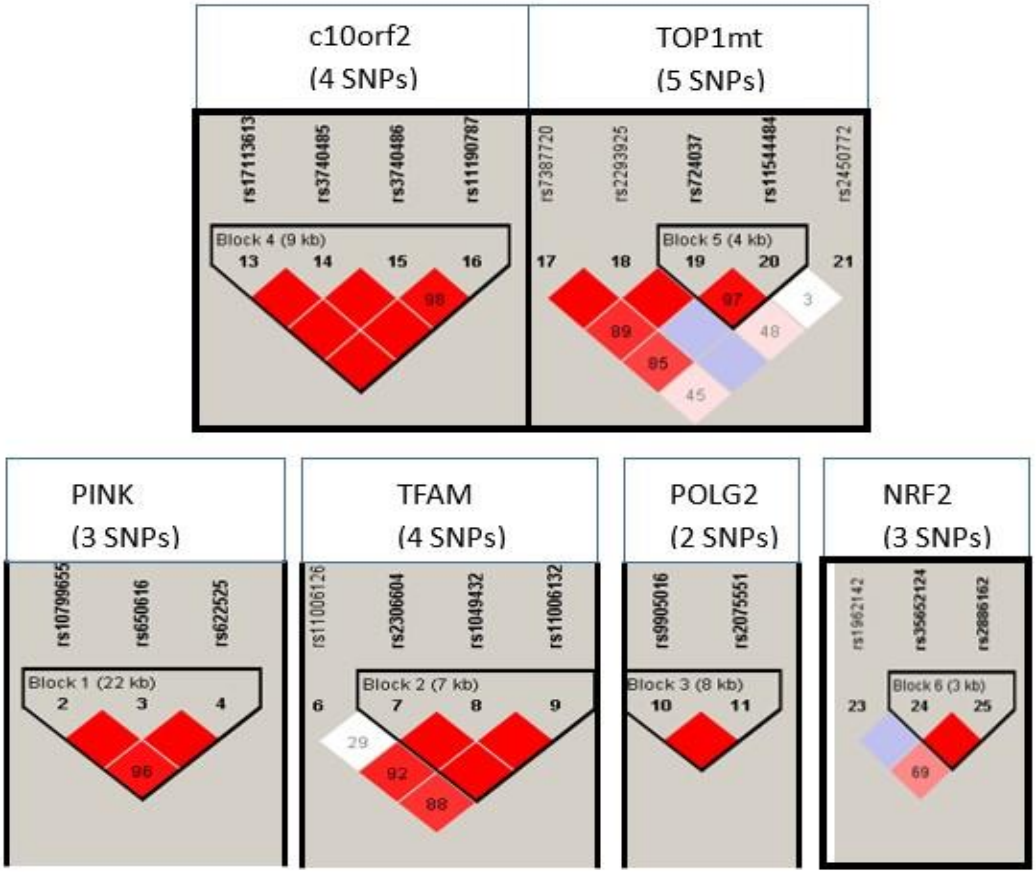


Figure 3.8: LD plot of the interrogated SNP markers (Pairwise LD).

The SNP ID is displayed along the top of the diagram. Colours represent D′ values (dark red=high inter-SNP D′; blue=statistically ambiguous D′; white=low inter-SNP D′), and r2 values are contained within blocks (bold=high r2).

3.8 Associations between alleles, genotypes, haplotypes and SN.

3.8.1 Univariate analysis of associations between genetic variation and SN

When we compared allele frequencies in cases with SN ever vs controls without SN ever we found that SNP rs1962142 in the *NRF2* gene was protective against HIVSN (chi squared P = 0.038 and pEMP1 = 0.029) (**Table 3.10**), while SNP rs724037 (*TOP1mt*) was significantly associated with the development of SN (chi - squared P = 0.047). SNP rs2856268 (*POLG*) was border line significantly associated with the development of SN (chi squared P = 0.07 and pEMP1 = 0.07). However, the associations of all three SNPs (rs1962142, rs724037 and rs2856268) were found to be insignificant after Bonferroni or pEMP2 correction for multiple testing.

When we compared genotype frequencies in cases with SN vs controls without SN (**Table 3.11**), we found that *NRF2* rs1962142 genotypes were significant for protective effect against SN development (Fishers P = 0.037, pEMP1=0.028) for both genotypic and dominant models. *POLG* rs2856268 genotypes were also significantly associated with SN (Fishers P = 0.04, pEMP1=0.04) for the dominant model. Both of these SNPs were therefore associated with SN in both allelic and genotypic models. Genotypes of the *TOP1mt* SNP rs724037 (that had significant allelic association with SN), were not significantly associated with SN for all three genotypic models (P > 0.05). All P-values became non-significant when Bonferroni or pEMP2 corrections were applied.

Table 3.10. Basic association test of alleles with the phenotype SN.

Gene	SNP ID	MAF in Cases	MAF Controls	OR	Chi-SQ P-value	Bonferroni P-value	EMP1	EMP2
PINK	rs10799655	0.29	0.30	0.94	0.750	1	0.760	1
	rs650616	0.43	0.41	1.05	0.797	1	0.833	1
	rs622525	0.16	0.20	0.79	0.315	1	0.338	0.999
CCDC19	rs2501325	0.13	0.14	0.89	0.646	1	0.703	1
ZFN648	rs7554182	0.28	0.23	1.29	0.221	1	0.239	0.995
NRF2	rs1962142	0.04	0.09	0.47	0.039*	1	0.029*	0.571
	rs35652124	0.17	0.19	0.90	0.637	1	0.633	1
	rs2886162	0.44	0.43	1.06	0.739	1	0.719	1
NFE2L3	rs4722585	0.28	0.23	1.26	0.268	1	0.323	0.998
TOP1mt	rs7387720	0.22	0.20	1.18	0.450	1	0.524	1
	rs2293925	0.02	0.01	1.64	0.565	1	0.691	1
	rs724037	0.24	0.16	1.59	0.047*	1	0.061	0.631
	rs11544484	0.45	0.40	1.21	0.300	1	0.323	0.999
	rs2450772	0.39	0.38	1.08	0.675	1	0.706	1

Gene	SNP ID	MAF in Cases	MAF Controls	OR	Chi-SQ P-value	Bonferroni P-value	EMP1	EMP2
TFAM	rs11006126	0.31	0.32	0.93	0.697	1	0.774	1
	rs2306604	0.19	0.22	0.79	0.295	1	0.314	0.999
	rs1049432	0.21	0.24	0.87	0.534	1	0.591	1
	rs11006132	0.22	0.24	0.88	0.552	1	0.563	1
C10orf2	rs17113613	0.15	0.13	1.16	0.578	1	0.597	1
	rs3740485	0.39	0.39	0.98	0.916	1	0.925	1
	rs3740486	0.39	0.38	1.01	0.962	1	1	1
C10orf2	rs11190787	0.38	0.38	0.99	0.965	1	0.981	1
POLG	rs2856268	0.16	0.11	1.62	0.073	1	0.074	0.793
LITAF	rs13333308	0.10	0.10	1.01	0.969	1	1	1
POLG2	rs9905016	0.17	0.16	1.09	0.726	1	0.776	1
	rs2075551	0.20	0.21	0.98	0.931	1	1	1

Significant P-value indicated by *

Table 3.11. Association of genotypes with SN using three genotype models.

GENE	SNP ID	Genotype Model	Genotype Counts in Cases	Genotype counts in Controls	Chi-Squared or Fisher's Exact Test P-value (Uncorrected)	P-value Bonferroni (Corrected)	EMP1	EMP2
PINK1		Genotypic	10/59/67	8/52/52	0.904	1	0.904	1
		Dominant	69/67	60/52	0.702	1	0.699	1
	rs10799655	Recessive	10/126	8/104	1	1	1	1
PINK1		Genotypic	22/71/42	20/52/39	0.679	1	0.676	1
		Dominant	93/42	72/39	0.586	1	0.538	1
	rs650616	Recessive	22/113	20/91	0.736	1	0.700	1
PINK1		Genotypic	5/34/97	2/40/70	0.151	1	0.152	0.964
		Dominant	39/97	42/70	0.174	1	0.173	0.963
	rs622525	Recessive	5/131	2/110	0.462	1	0.462	0.999
CCDC19		Genotypic	4/27/105	3/26/83	0.833	1	0.832	1
		Dominant	31/105	29/83	0.655	1	0.655	1
	rs2501325	Recessive	4/132	3/109	1	1	1	1
ZNF648		Genotypic	9/57/70	4/43/65	0.426	1	0.429	1
		Dominant	66/70	47/65	0.309	1	0.308	0.999
	rs7554182	Recessive	9/127	4/108	0.393	1	0.394	0.999
NRF2		Genotypic	0/12/124	0/20/91	0.037*	1	0.028*	0.548
		Dominant	12/124	20/91	0.037*	1	0.029*	0.496
	rs1962142	Recessive	0/136	0/111	1	1	1	1
NRF2		Genotypic	5/37/94	2/38/71	0.430	1	0.425	1
		Dominant	42/94	40/71	0.417	1	0.376	1
	rs35652124	Recessive	5/131	2/109	0.464	1	0.463	0.999
NRF2	rs2886162	Genotypic	30/60/45	15/61/30	0.102	1	0.102	0.892
		Dominant	90/45	76/30	0.484	1	0.402	1
		Recessive	30/105	15/91	0.134	1	0.102	0.884

GENE	SNP ID	Genotype Model	Genotype Counts in Cases	Genotype counts in Controls	Chi-Squared or Fisher's Exact Test P-value (Uncorrected)	P-value Bonferroni (Corrected)	EMP1	EMP2
NRF3/NFE 2L3		Genotypic	13/49/74	7/38/67	0.546	1	0.547	1
		Dominant	62/74	45/67	0.440	1	0.440	1
	rs4722585	Recessive	13/123	7/105	0.362	1	0.362	0.999
TOP1MT		Genotypic	9/43/84	5/34/73	0.767	1	0.768	1
		Dominant	52/84	39/73	0.5988	1	0.5976	1
	rs7387720	Recessive	9/127	5/107	0.585	1	0.584	1
TOP1MT		Genotypic	0/4/132	0/2/109	0.693	1	0.691	1
		Dominant	4/132	2/109	0.693	1	0.692	1
	rs2293925	Recessive	0/136	0/111	1	1	1	1
TOP1MT		Genotypic	10/43/81	4/27/77	0.170	1	0.169	0.977
		Dominant	53/81	31/77	0.103	1	0.083**	0.858
	rs724037	Recessive	10/124	4/104	0.273	1	0.263	0.991
TOP1MT		Genotypic	28/65/42	19/52/41	0.596	1	0.595	1
		Dominant	93/42	71/41	0.417	1	0.383	1
	rs11544484	Recessive	28/107	19/93	0.516		0.470	1
TOP1MT		Genotypic	19/69/48	15/54/43	0.893	1	0.892	1
		Dominant	88/48	69/43	0.692	1	0.691	1
	rs2450772	Recessive	19/117	15/97	1	1	1	1
TFAM		Genotypic	16/51/69	10/52/50	0.357	1	0.355	0.999
		Dominant	67/69	62/50	0.373	1	0.370	0.999
	rs11006126	Recessive	16/120	10/102	0.536	1	0.534	1
TFAM		Genotypic	5/40/90	5/40/67	0.528	1	0.526	1
	rs2306604	Dominant	45/90	45/67	0.290	1	0.292	0.998

GENE	SNP ID	Genotype Model	Genotype Counts in Cases	Genotype counts in Controls	Chi-Squared or Fisher's Exact Test P-value (Uncorrected)	P-value Bonferroni (Corrected)	EMP1	EMP2
TFAM		Recessive	5/130	5/107	0.759	1	0.758	1
TFAM	rs1049432	Genotypic	8/42/86	5/43/64	0.459	1	0.462	1
		Dominant	50/86	48/64	0.362	1	0.361	0.999
		Recessive	8/128	5/107	0.777	1	0.780	1
TFAM		Genotypic	9/42/85	6/42/63	0.524	1	0.522	1
		Dominant	51/85	48/63	0.365	1	0.362	0.999
	rs11006132	Recessive	9/127	6/105	0.793	1	0.794	1
C10orf2		Genotypic	3/34/98	1/27/83	0.799	1	0.781	1
		Dominant	37/98	28/83	0.772	1	0.751	1
	rs17113613	Recessive	3/132	1/110	0.629	1	0.567	1
C10orf2		Genotypic	19/65/49	18/51/42	0.867	1	0.857	1
		Dominant	84/49	69/42	0.895	1	0.896	1
	rs3740485	Recessive	19/114	18/93	0.722	1	0.708	1
C10orf2		Genotypic	18/69/49	18/50/44	0.611	1	0.611	1
		Dominant	87/49	68/44	0.601	1	0.603	1
	rs3740486	Recessive	18/118	18/94	0.589	1	0.587	1
C10orf2		Genotypic	19/62/51	17/49/43	0.953	1	0.943	1
		Dominant	81/51	66/43	1	1	0.996	1
	rs11190787	Recessive	19/113	17/92	0.857	1	0.813	1
POLG		Genotypic	2/40/93	2/20/90	0.087**	1	0.086**	0.839
		Dominant	42/93	22/90	0.043*	1	0.042*	0.569
	rs2856268	Recessive	2/133	2/110	1	1	0.834	1

GENE	SNP ID	Genotype Model	Genotype Counts in Cases	Genotype counts in Controls	Chi-Squared or Fisher's Exact Test P-value (Uncorrected)	P-value Bonferroni (Corrected)	EMP1	EMP2
LITAF	rs13333308	Genotypic	1/25/110	0/22/90	0.930	1	0.930	1
		Dominant	26/110	22/90	1	1	1	1
		Recessive	1/135	0/112	1	1	1	1
POLG2	rs9905016	Genotypic	6/35/94	5/26/80	0.943	1	0.930	1
		Dominant	41/94	31/80	0.778	1	0.725	1
		Recessive	6/129	5/106	1	1	0.927	1
POLG2	rs2075551	Genotypic	7/41/88	6/34/72	1	1	1	1
		Dominant	48/88	40/72	1	1	1	1
		Recessive	7/129	6/106	1	1	1	1

Significant P-value indicated by *

The haplotype frequencies in the SN positive and SN negative group were compared. Asymptotic P-values were also generated (**Table 3.12**), for all haplotypes. Two haplotypes were significantly associated with SN (**Table 3.12**). The haplotype rs7387720|rs2293925|rs724037 GGT in *TOP1mt* conferred a protective effect against the development of SN (P = 0.004). This haplotype included rs724037 that was associated with SN in an allelic model. Another haplotype in *TOP1mt* also containing this SNP, rs2293925|rs724037|rs11544484 GGC, was weakly associated with SN (P = 0.08 on Table 3.12). The TFAM rs11006126|rs2306604|rs1049432 CGG haplotype was significantly associated with the development of SN (P = 0.021). None of the SNPs in this haplotype were independently associated with SN risk in allelic or genotypic mode

Table 3.12 Haplotype association (Univariate) with SN.

LOCUS	HAPLOTYPE	CASES	CONTROLS	P-value	pEMP1	pEMP2	SNPS	GENES
WIN1	TGA	0.158	0.192	0.322	0.333	1	rs10799655|rs650616|rs622525	*PINK*
WIN1	CAT	0.429	0.418	0.811	0.817	1	rs10799655|rs650616|rs622525	*PINK*
WIN1	TGT	0.133	0.113	0.494	0.489	1	rs10799655|rs650616|rs622525	*PINK*
WIN1	CGT	0.279	0.277	0.945	0.944	1	rs10799655|rs650616|rs622525	*PINK*
WIN2	GAC	0.009	0.012	0.764	0.661	1	rs650616|rs622525|rs2501325	*PINK – CCDC19*
WIN2	ATC	0.047	0.058	0.567	0.497	1	rs650616|rs622525|rs2501325	*PINK – CCDC19*
WIN2	GTC	0.073	0.073	0.999	0.998	1	rs650616|rs622525|rs2501325	*PINK – CCDC19*
WIN2	GAA	0.153	0.185	0.343	0.340	1	rs650616|rs622525|rs2501325	*PINK – CCDC19*
WIN2	ATA	0.380	0.357	0.602	0.582	1	rs650616|rs622525|rs2501325	*PINK – CCDC19*
WIN2	GTA	0.339	0.315	0.581	0.582	1	rs650616|rs622525|rs2501325	*PINK – CCDC19*
WIN3	TCT	0.032	0.017	0.308	0.168	0.999	rs622525|rs2501325|rs7554182	*PINK – CCDC19-ZFN*
WIN3	AAT	0.037	0.043	0.752	0.687	1	rs622525|rs2501325|rs7554182	*PINK – CCDC19-ZFN*
WIN3	TAT	0.201	0.159	0.232	0.191	0.999	rs622525|rs2501325|rs7554182	*PINK – CCDC19-ZFN*
WIN3	TCC	0.090	0.116	0.337	0.319	1	rs622525|rs2501325|rs7554182	*PINK – CCDC19-ZFN*

Locus	Haplotype	Cases	Controls	P-value	pEMP1	pEMP2	SNPS	GENES
WIN3	AAC	0.118	0.145	0.374	0.381	1	rs622525|rs2501325|rs7554182	*PINK – CCDC19-ZFN*
WIN3	TAC	0.523	0.520	0.951	0.922	1	rs622525|rs2501325|rs7554182	*PINK – CCDC19-ZFN*
WIN4	ATA	0.040	0.071	0.133	0.126	0.991	rs1962142|rs35652124|rs2886162	*NRF2*
WIN4	GTA	0.405	0.368	0.415	0.329	1	rs1962142|rs35652124|rs2886162	*NRF2*
WIN4	GCG	0.174	0.189	0.669	0.711	1	rs1962142|rs35652124|rs2886162	*NRF2*
WIN4	GTG	0.382	0.372	0.824	0.682	1	rs1962142|rs35652124|rs2886162	*NRF2*
WIN6	GAG	0.013	0.009	0.658	0.510	1	rs7387720|rs2293925|rs724037	*TOP1mt*
WIN6	GGG	0.207	0.150	0.101	0.102	0.988	rs7387720|rs2293925|rs724037	*TOP1mt*
WIN6	AGG	0.019	0.014	0.674	0.689	1	rs7387720|rs2293925|rs724037	*TOP1mt*
WIN6	GGT	0.003	0.037	0.004*	0.012*	0.689	rs7387720|rs2293925|rs724037	*TOP1mt*
WIN6	AGT	0.758	0.790	0.406	0.421	1	rs7387720|rs2293925|rs724037	*TOP1mt*
WIN7	AGC	0.015	0.009	0.562	0.425	1	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN7	GGC	0.220	0.158	0.080	0.086	0.958	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN7	GTC	0.215	0.237	0.557	0.577	1	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN7	GTT	0.550	0.596	0.304	0.337	1	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN8	GCA	0.155	0.122	0.297	0.295	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TCA	0.021	0.009	0.272	0.149	0.999	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TTA	0.218	0.242	0.531	0.518	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	GCG	0.081	0.047	0.129	0.087	0.980	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TCG	0.192	0.226	0.356	0.382	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TTG	0.333	0.355	0.619	0.626	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN9	CGT	0.198	0.232	0.362	0.331	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	TGT	0.015	0.005	0.265	0.229	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	CAG	0.041	0.067	0.190	0.160	0.998	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	TAG	0.145	0.156	0.726	0.711	1	rs11006126|rs2306604|rs1049432	*TFAM*

LOCUS	HAPLOTYPE	CASES	CONTROLS	P- value	pEMP1	pEMP2	SNPS	GENES
WIN9	CGG	0.066	0.022	0.021*	0.020*	0.633	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	TGG	0.535	0.518	0.701	0.681	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN10	GTG	0.215	0.239	0.527	0.556	1	rs2306604|rs1049432|rs11006132	*TFAM*
WIN10	AGA	0.187	0.225	0.297	0.293	1	rs2306604|rs1049432|rs11006132	*TFAM*
WIN10	GGA	0.598	0.536	0.167	0.173	0.998	rs2306604|rs1049432|rs11006132	*TFAM*
WIN11	TGA	0.034	0.039	0.737	0.616	1	rs1049432|rs11006132|rs17113613	*TFAM-C10orf2*
WIN11	GAA	0.114	0.093	0.442	0.400	1	rs1049432|rs11006132|rs17113613	*TFAM-C10orf2*
WIN11	TGG	0.181	0.200	0.607	0.584	1	rs1049432|rs11006132|rs17113613	*TFAM-C10orf2*
WIN11	GAG	0.671	0.669	0.951	0.924	1	rs1049432|rs11006132|rs17113613	*TFAM-C10orf2*
WIN12	GAC	0.035	0.036	0.916	0.890	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	AAC	0.114	0.095	0.491	0.466	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	GGC	0.057	0.069	0.578	0.585	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	AGC	0.188	0.194	0.857	0.834	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	GGT	0.129	0.138	0.777	0.777	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	AGT	0.478	0.468	0.826	0.798	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN13	ACT	0.149	0.132	0.597	0.566	1	rs17113613|rs3740485|rs3740486	*C10orf2*
WIN13	GCT	0.239	0.255	0.670	0.714	1	rs17113613|rs3740485|rs3740486	*C10orf2*
WIN13	GTC	0.613	0.613	0.999	0.943	1	rs17113613|rs3740485|rs3740486	*C10orf2*
WIN14	CTG	0.383	0.385	0.969	0.949	1	rs3740485|rs3740486|rs11190787	*C10orf2*
WIN14	TCT	0.617	0.616	0.969	0.943	1	rs3740485|rs3740486|rs11190787	*C10orf2*
WIN17	TC	0.174	0.162	0.726	0.770	1	rs9905016|rs2075551	*POLG2*
WIN17	CC	0.026	0.041	0.363	0.337	1	rs9905016|rs2075551	*POLG2*
WIN17	CG	0.800	0.797	0.941	0.935	1	rs9905016|rs2075551	*POLG2*

P-value (Significant) shown by *

3.8.2 Multivariate analysis of association with SN

Multivariate analysis was performed only on SNPs, genotypes and haplotypes with P-values <0.5 during univariate analysis of association with SN. The multivariate analysis involved inclusion of covariates that were significant during analyses of demographic data i.e. age and height. These covariates were tested as to whether they had any additive effect on the development of SN. Logistic regressions (OR) results are shown for categorical data (SN).

All the alleles and genotypes that were found to be significantly associated with SN in univariate analysis were found to be non-significant when tested using multivariate analysis (**Table 3.13 and Table 3.14)**.

Table 3.13. Logistic regression analysis of association between alleles and SN including age and height as covariates.

Gene	SNP	P-Value	Bonferroni P-value	EMP1	EMP2	OR
NRF2	rs1962142	0.137	1	0.134	0.947	0.543
TOP1mt	rs724037	0.083	1	0.080	0.835	1.510
POLG	rs2856268	0.244	1	0.226	0.999	1.413

Genotypic analysis including the covariates yielded no significant P-values for all SNPs (rs1962142, rs724037, rs2856268 & P = NA, 0.22, 0.32) respectively (**Table 3.14**). The rs1962142 SNP generated a not available (NA) result when interrogated for any additive effect to the phenotype of SN.

Table 3.14: Logistic regression analysis of association between genotypes and SN including age and height as covariates

Gene	SNP	Genotypic model	P-Value	Bonferroni P-value	EMP1	EMP2	OR
NRF2	rs1962142	Genotypic	NA	1	1	1	NA
NRF2	rs1962142	Dominant	0.137	1	0.125	0.948	0.543
POLG	rs2856268	Genotypic	0.323	1	0.578	1	2.263
POLG	rs2856268	Dominant	0.440	1	0.184	0.989	0.792

The three haplotypes that were significantly associated with SN in univariate analysis in **Table 3.12** were tested for association using the multivariate model (using age and height as cofactors) and two maintained their significance (**Table 3.15**). Empirical P-value (pEMP1) was also computed for the significant haplotypes using hundred thousand (100 000) permutations. The haplotypes maintained their significance (**Table 3.15**). The haplotype GGT had a significant pEMP1 of 0.007 and haplotype CGG had a significant pEMP1 of 0.017. The *TOP1mt* haplotype GGC which was borderline significant during univariate analysis was found to be non-significant during multivariate analysis. All haplotypes were non-significant after correction for pEMP2.

Table 3.15. Additive effect of age and height to haplotype association with SN (Logistic regression analysis).

GENE	HAPLOTYPE	FREQUENCY	OR	STAT	P-VALUE	EMP1	EMP2
TOP1mt	rs7387720|rs2293925|rs724037 **(GGT)**	0.019*	0.04	4.44	0.035*	0.007*	0.703
TOP1mt	rs2293925|rs724037|rs11544484 **(GGC)**	0.191	1.47	2.5	0.114	0.114	0.987
TFAM	rs11006126|rs2306604|rs1049432 **(CGG)**	0.047 *	3.82	4.87	0.027 *	0.017*	0.609

Significant P-value indicated by *

A summary of all the alleles, genotypes and haplotypes that were significantly (P<0.05) associated in univariate analysis with SN, and results after multivariate analysis is shown in **Table 3.15** below.

Table 3.16: Summary of results from association of genetic variants with SN.

				Univariate analyses			Multivariate analyses			
Gene	**SNP**	**Model**	**Associated with**	**P**	**pEMP1**	**OR**	**P-value**	**pEMP1**	**pEMP2**	**OR**
NRF2	rs1962142	Allelic	SN	0.039*	0.029 *	0.47	0.137	0.134	0.947	0.54
NRF2	rs1962142	Dominant	SN	0.037*	0.028*	-	0.137	0.125	0.948	0.54
NRF2	rs1962142	Genotypic	SN	0.037*	0.029*	-	NA	1	1	NA
POLG	rs2856268	Dominant	SN	0.043*	0.041*	-	0.440	0.184	0.989	0.792
TOP1mt	rs724037	Allelic	SN	0.047*	0.061	1.62	0.083	0.080	0.835	1.51
TOP1mt	rs7387720|rs2293925|rs724037	Haplotype **(GGT)**	SN	0.004*	0.012*	0.04	0.035 *	0.007*	0.703	0.04
TFAM	rs11006126|rs2306604|rs1049432	Haplotype **(CGG)**	SN	0.021*	0.020*	3.82	0.027 *	0.017*	0.609	3.82

Significant P-value indicated by *

3.9 Associations between alleles, genotype, haplotypes and pain intensity

3.9.1 Univariate Associations between genetic variants and pain intensity

When we compared pain intensity phenotype to SNP risk allele, we found that *NRF3* gene SNP (rs4722585) was significantly associated with increased pain intensity (T-test P = 0.014 and pEMP1 = 0.014) (**Table 3.16)**. The rs7387720 SNP of *TOP1mt* gene was border-line significant for association with decreased pain intensity (T-test P = 0.052 and pEMP1 = 0.053) (**Table 3.16).** However, these associations were not significant after adjustments for multiple testing.

When we performed genotype association with pain intensity (**Table 3.17**) we found that the *NRF* SNP rs4722585 was significantly associated with an increase in pain intensity (T-test P = 0.045; pEMP1 = 0.059) and (T-test P = 0.018; pEMP1 = 0.023) for both the genotypic and dominant models, respectively. The rs7387720 SNP in *TOP1mt* was also significant for protective effect against pain intensity (T-test P = 0.030; pEMP1 = 0.033) for the dominant model. All P-values became non-significant when Bonferroni or pEMP2 corrections were applied. SNPs in blue in the Table below had the lowest P-values for association with pain intensity in the cohort but all were above significance level of 0.05.

Table 3.17. Basic association test of alleles with the phenotype pain intensity.

Gene	SNP ID	Beta	T-test P-value	Bonferroni P-value	pEMP1	pEMP2
PINK	rs10799655	0.141	0.101	1	0.099	0.886
	rs650616	-0.119	0.171	1	0.171	0.977
	rs622525	0.115	0.182	1	0.184	0.982
CCDC19	rs2501325	0.056	0.517	1	0.532	1
ZNF648	rs7554182	-0.032	0.710	1	0.717	1
NRF2	rs1962142	-0.066	0.449	1	0.465	1
	rs35652124	0.091	0.295	1	0.300	0.999
	rs2886162	0.036	0.678	1	0.676	1.000
NRF3	rs4722585	0.209	0.014*	0.377	0.014*	0.260
TOP1mt	rs7387720	-0.166	0.053**	1	0.053	0.673
	rs2293925	-0.084	0.330	1	0.352	1.000
	rs724037	-0.146	0.093	1	0.093	0.863
	rs11544484	-0.105	0.226	1	0.225	0.994
	rs2450772	0.073	0.397	1	0.405	1
TFAM	rs11006126	0.076	0.382	1	0.390	1
	rs2306604	0.018	0.839	1	0.838	1
	rs1049432	0.051	0.558	1	0.563	1
	rs11006132	0.019	0.831	1	0.835	1
C10orf2	rs17113613	-0.103	0.233	1	0.234	0.995
	rs3740485	-0.061	0.486	1	0.483	1
	rs3740486	-0.076	0.381	1	0.378	1
C10orf2	rs11190787	-0.050	0.568	1	0.567	1
POLG	rs2856268	-0.020	0.820	1	0.820	1
LITAF	rs13333308	0.015	0.861	1	0.868	1
POLG2	rs9905016	0.085	0.326	1	0.329	1
	rs2075551	0.118	0.171	1	0.172	0.977

P-value (Significant) : *
P-value (border line Significant) : **

Table 3.18. Association of genotypes with pain intensity using three models of genotype inheritance.

GENE	SNP ID	Genotype Model	Beta	STAT	T-test P-Value	P-value Bonferroni	EMP1	EMP2
PINK1	rs10799655	Genotypic	NA	5.67	0.715	1	0.727	1
		Dominant	0.188	2.209	0.029	0.750	0.031	0.499
		Recessive	-0.021	-0.242	0.809	1	0.810	1
PINK1	rs650616	Genotypic	NA	3.37	0.185	1	0.112	0.893
		Dominant	-0.046	-0.534	0.594	1	0.598	1
		Recessive	-0.158	-1.843	0.068**	1	0.057	0.715
PINK1	rs622525	Genotypic	NA	2.07	0.355	1	0.605	1
		Dominant	0.124	1.443	0.151	1	0.148	0.973
		Recessive	0.033	0.388	0.699	1	0.719	1
CCDC19	rs2501325	Genotypic	NA	0.70	0.706	1	0.398	1
		Dominant	0.039	0.447	0.656	1	0.659	1
		Recessive	0.070	0.814	0.417	1	0.450	1
ZFN648	rs7554182	Genotypic	NA	0.22	0.894	1	0.659	1
		Dominant	-0.020	-0.230	0.818	1	0.802	1
		Recessive	-0.040	-0.463	0.645	1	0.638	1
NRF2	rs1962142	Genotypic	NA	NA	NA	1	1	1
		Dominant	-0.066	-0.76	0.449	1	0.429	1
		Recessive	NA	NA	NA	1	1	1
NRF2	rs35652124	Genotypic	NA	2.31	0.316	1	0.149	0.950
		Dominant	0.055	0.633	0.528	1	0.547	1
		Recessive	0.129	1.510	0.133	1	0.109	0.934
NRF2	rs2886162	Genotypic	NA	0.59	0.745	1	0.609	1
		Dominant	0.001	0.017	0.987	1	0.985	1

P-value (Significant) : *

P-value (border line Significant) : **

GENE	SNP ID	Genotype Model	Beta	STAT	T-test P-Value	P-value Bonferroni	EMP1	EMP2
		Recessive	0.062	0.719	0.473	1	0.472	1
NRF3		Genotypic	NA	6.20	0.045	1	0.059	0.618
		Dominant	0.202	2.386	0.018	0.479	0.023	0.365
	rs4722585	Recessive	0.129	1.510	0.133	1	0.150	0.936
TOP1MT		Genotypic	NA	4.82	0.090**	1	0.357	1
		Dominant	-0.186	-2.195	0.030	0.776	0.033	0.511
	rs7387720	Recessive	-0.048	-0.556	0.579	1	0.577	1
TOP1MT		Genotypic	NA	NA	NA	1	1	1
		Dominant	-0.084	-0.978	0.330	1	0.326	1
	rs2293925	Recessive	NA	NA	NA	1	1	1
TOP1MT		Genotypic	NA	3.16	0.206	1	0.303	1
		Dominant	-0.153	-1.777	0.078**	1	0.076	0.835
	rs724037	Recessive	-0.066	-0.757	0.451	1	0.458	1
TOP1MT		Genotypic	NA	2.57	0.277	1	0.178	0.981
		Dominant	-0.041	-0.469	0.640	1	0.613	1
	rs11544484	Recessive	-0.138	-1.606	0.111	1	0.097	0.884
TOP1MT		Genotypic	NA	3.60	0.165	1	0.165	0.980
		Dominant	-0.008	-0.094	0.925	1	0.931	1
	rs2450772	Recessive	0.152	1.784	0.077**	1	0.070	0.774
TFAM		Genotypic	NA	0.83	0.662	1	0.503	1
		Dominant	0.077	0.898	0.371	1	0.412	1
	rs11006126	Recessive	0.041	0.479	0.632	1	0.636	1
TFAM		Genotypic	NA	0.21	0.899	1	0.866	1
	rs2306604	Dominant	0.028	0.329	0.743	1	0.727	1

P-value (Significant) : *
P-value (border line Significant) : **

GENE	SNP ID	Genotype Model	Beta	STAT	T-test P-Value	P-value Bonferroni	EMP1	EMP2
TFAM	rs2306604	Recessive	-0.019	-0.223	0.824	1	0.839	1
TFAM	rs1049432	Genotypic	NA	0.94	0.624	1	0.359	1
		Dominant	0.022	0.260	0.795	1	0.803	1
		Recessive	0.084	0.973	0.332	1	0.358	1
TFAM	rs11006132	Genotypic	NA	0.24	0.887	1	0.685	1
		Dominant	0.003	0.030	0.976	1	0.978	1
		Recessive	0.041	0.472	0.638	1	0.630	1
C10orf2	rs17113613	Genotypic	NA	3.90	0.142	1	0.144	0.603
		Dominant	-0.061	-0.704	0.483	1	0.501	1
		Recessive	-0.168	-1.967	0.051**	1	0.057	0.603
C10orf2	rs3740485	Genotypic	NA	0.82	0.663	1	0.402	1
		Dominant	-0.028	-0.325	0.746	1	0.771	1
		Recessive	-0.079	-0.909	0.365	1	0.354	1
C10orf2	rs3740486	Genotypic	NA	0.98	0.612	1	0.347	1
		Dominant	-0.047	-0.541	0.5896	1	0.611	1
		Recessive	-0.082	-0.955	0.341	1	0.331	1
C10orf2	rs11190787	Genotypic	NA	0.52	0.770	1	0.504	1
		Dominant	-0.025	-0.285	0.7762	1	0.786	1
		Recessive	-0.063	-0.725	0.470	1	0.467	1
POLG	rs2856268	Genotypic	NA	1.14	0.565	1	0.452	1
		Dominant	-0.041	-0.470	0.639	1	0.646	1
		Recessive	0.074	0.859	0.392	1	0.433	1

GENE	SNP ID	Genotype Model	Beta	STAT	T-test P-Value	P-value Bonferroni	EMP1	EMP2
LITAF	rs13333308	Genotypic	NA	0.03	0.985	1	0.966	1
		Dominant	0.015	0.173	0.8631	1	0.891	1
		Recessive	0.005	0.062	0.951	1	0.988	1
POLG2	rs9905016	Genotypic	NA	1.70	0.427	1	0.196	0.985
		Dominant	0.054	0.628	0.531	1	0.475	1
		Recessive	0.111	1.290	0.199	1	0.203	0.989
POLG2	rs2075551	Genotypic	NA	3.21	0.201	1	0.074	0.768
		Dominant	0.075	0.874	0.384	1	0.351	1
		Recessive	0.151	1.764	0.080**	1	0.080	0.786

P-value (Significant) : *

P-value (border line Significant) : **

When we compared pain intensity to haplotypes, we found that there was no significant association with all haplotypes (**Table 3.19**). All computed P-values were ($P > 0.05$). The haplotypes in blue in the table below had the lowest P-values (< 0.1 but > 0.05).

Table 3.19. Haplotype Association with pain intensity.

LOCUS	HAPLOTYPE	Beta	R2	STAT	P-value	pEMP1	pEMP2	SNPS	GENES
WIN1	TGA	0.858	0.016	1.466	0.145	0.157	1	rs10799655|rs650616|rs622525	*PINK*
WIN1	CAT	-0.646	0.014	-1.377	0.171	0.165	1	rs10799655|rs650616|rs622525	*PINK*
WIN1	TGT	0.341	0.002	0.509	0.612	0.604	1	rs10799655|rs650616|rs622525	*PINK*
WIN1	CGT	-0.034	0.000	-0.068	0.946	0.954	1	rs10799655|rs650616|rs622525	*PINK*
WIN2	ATC	-0.831	0.004	-0.695	0.489	0.502	1	rs650616|rs622525|rs2501325	*PINK – CCDC19*
WIN2	GTC	1.241	0.014	1.395	0.165	0.177	1	rs650616|rs622525|rs2501325	*PINK – CCDC19*
WIN2	GAA	0.803	0.013	1.343	0.182	0.193	1	rs650616|rs622525|rs2501325	*PINK – CCDC19*
WIN2	ATA	-0.576	0.010	-1.162	0.247	0.244	1	rs650616|rs622525|rs2501325	*PINK – CCDC19*
WIN2	GTA	-0.226	0.002	-0.456	0.649	0.637	1	rs650616|rs622525|rs2501325	*PINK – CCDC19*
WIN3	TCT	1.762	0.011	1.205	0.230	0.239	1	rs622525|rs2501325|rs7554182	*PINK – CCDC19*
WIN3	AAT	-0.896	0.002	-0.543	0.588	0.598	1	rs622525|rs2501325|rs7554182	*PINK – CCDC19-ZFN*
WIN3	TAT	-0.433	0.004	-0.714	0.476	0.482	1	rs622525|rs2501325|rs7554182	*PINK – CCDC19-ZFN*
WIN3	TCC	0.182	0.000	0.210	0.834	0.843	1	rs622525|rs2501325|rs7554182	*PINK – CCDC19-ZFN*
WIN3	AAC	1.249	0.025	1.837	0.068**	0.075	0.928	rs622525|rs2501325|rs7554182	*PINK – CCDC19-ZFN*
WIN3	TAC	-0.432	0.007	-0.969	0.334	0.331	1	rs622525|rs2501325|rs7554182	*PINK – CCDC19-ZFN*
WIN4	ATA	-0.847	0.004	-0.760	0.449	0.449	1	rs1962142|rs35652124|rs2886162	*PINK – CCDC19-ZFN*
WIN4	GTA	0.337	0.004	0.775	0.440	0.456	1	rs1962142|rs35652124|rs2886162	*NRF2*
WIN4	GCG	0.606	0.008	1.052	0.295	0.324	1	rs1962142|rs35652124|rs2886162	*NRF2*
WIN4	GTG	-0.661	0.014	-1.396	0.165	0.174	1	rs1962142|rs35652124|rs2886162	*NRF2*
WIN6	GAG	-2.153	0.007	-1.002	0.318	0.303	1	rs7387720|rs2293925|rs724037	*NRF2*
WIN6	GGG	-0.934	0.022	-1.756	0.081**	0.078	0.959	rs7387720|rs2293925|rs724037	*TOP1mt*
WIN6	AGG	1.288	0.004	0.767	0.445	0.431	1	rs7387720|rs2293925|rs724037	*TOP1mt*

P-value (Significant) : *
P-value (border line Significant) : **

LOCUS	HAPLOTYPE	Beta	R2	STAT	P-value	pEMP1	pEMP2	SNPS	GENES
WIN6	AGT	0.801	0.020	1.640	0.103	0.105	0.982	rs7387720|rs2293925|rs724037	*TOP1mt*
WIN7	AGC	-1.826	0.007	-0.978	0.330	0.314	1	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN7	GGC	-0.751	0.016	-1.463	0.146	0.141	1	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN7	GTC	0.158	0.001	0.304	0.762	0.757	1	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN7	GTT	0.545	0.011	1.229	0.221	0.212	1	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN8	GCA	-0.705	0.008	-1.053	0.294	0.287	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TCA	2.068	0.010	1.140	0.257	0.244	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TTA	1.057	0.022	1.716	0.089**	0.096	0.968	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	GCG	-1.289	0.017	-1.516	0.132	0.126	0.997	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TCG	-0.012	0.000	-0.023	0.982	0.976	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TTG	0.002	0.000	0.003	0.997	0.996	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN9	CGT	0.542	0.008	1.022	0.309	0.312	1	rs11006126|rs2306604|rs1049432	*TOP1mt*
WIN9	TGT	-2.805	0.017	-1.505	0.135	0.159	0.998	rs11006126|rs2306604|rs1049432	*TOP1mt*
WIN9	CAG	0.329	0.000	0.222	0.825	0.846	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	TAG	0.086	0.000	0.135	0.893	0.898	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	CGG	-0.118	0.000	-0.137	0.891	0.889	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	TGG	-0.249	0.002	-0.573	0.567	0.582	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN10	GTG	0.309	0.003	0.588	0.558	0.549	1	rs2306604|rs1049432|rs11006132	*TFAM*
WIN10	AGA	0.120	0.000	0.209	0.834	0.853	1	rs2306604|rs1049432|rs11006132	*TFAM*
WIN10	GGA	-0.151	0.001	-0.344	0.732	0.733	1	rs2306604|rs1049432|rs11006132	*TFAM*
WIN11	TGA	-0.964	0.003	-0.598	0.551	0.575	1	rs1049432|rs11006132|rs17113613	*TFAM*
WIN11	GAA	-0.745	0.007	-0.966	0.336	0.346	1	rs1049432|rs11006132|rs17113613	*TFAM*
WIN11	TGG	0.474	0.005	0.842	0.402	0.402	1	rs1049432|rs11006132|rs17113613	*TFAM-C10orf2*
WIN11	GAG	0.181	0.001	0.388	0.699	0.679	1	rs1049432|rs11006132|rs17113613	*TFAM-C10orf2*
WIN12	GAC	-1.358	0.006	-0.865	0.389	0.407	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	AAC	-0.802	0.008	-1.042	0.299	0.313	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	GGC	-0.451	0.001	-0.432	0.666	0.642	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	AGC	0.149	0.000	0.235	0.815	0.799	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*

LOCUS	HAPLOTYPE	Beta	R2	STAT	P-value	pEMP1	pEMP2	SNPs	GENES
WIN12	GGT	0.631	0.006	0.936	0.351	0.383	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	AGT	0.124	0.001	0.263	0.793	0.799	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN13	ACT	-0.828	0.013	-1.310	0.193	0.197	1	rs17113613|rs3740485|rs3740486	*TFAM-C10orf2*
WIN13	GCT	0.057	0.000	0.109	0.914	0.918	1	rs17113613|rs3740485|rs3740486	*TFAM-C10orf2*
WIN13	GTC	0.374	0.005	0.795	0.428	0.413	1	rs17113613|rs3740485|rs3740486	*C10orf2*
WIN14	CTG	-0.289	0.003	-0.613	0.541	0.533	1	rs3740485|rs3740486|rs11190787	*C10orf2*
WIN14	TCT	0.372	0.005	0.791	0.430	0.421	1	rs3740485|rs3740486|rs11190787	*C10orf2*
WIN17	TC	0.558	0.007	0.987	0.326	0.337	1	rs9905016|rs2075551	*POLG2*
WIN17	CC	1.598	0.009	1.116	0.267	0.227	1	rs9905016|rs2075551	*POLG2*
WIN17	CG	-0.741	0.014	-1.372	0.172	0.188	1	rs9905016|rs2075551	*POLG2*

P-value (Significant) : *

P-value (border line Significant) : **

3.9.2 Multivariate analysis of genetic associations with pain intensity

Multivariate analysis was performed only on SNPs, genotypes and haplotypes with P-values < 0.5 during univariate analysis. The multivariate analysis involved inclusion of covariates that were significantly associated with SN during analyses of demographic data i.e. age and height. These covariates were tested as to whether they had any additive effect on pain intensity

The two SNPs *NRF3 (*rs4722585) and *TOP1mt* (rs7387720) associated with pain intensity in SN patients in univariate analysis, were further analysed for association using the multivariate model (linear regression analysis) and they maintained their significance when age and height were also considered (**Table 3.20**). The rs4722585 SNP had P = 0.013 for association with an increase in pain, and a significant EMP1 of 0.013 was also obtained. TOP1mt SNP rs7387720 was border line significant P = 0.054 for association with a decrease in pain intensity in SN patients and was also border line significant after correction EMP1 of 0.054. However, pEMP2 values were all above 0.05, indicating that these alleles were not significantly associated with pain intensity after correction for multiple testing. Similarly, genotypes shown in **Table 3.21** were not significantly associated with pain intensity after correction for multiple testing.

Table 3.20. Linear regression analysis of alleles in association with pain intensity including age and height as covariates.

Gene	SNP	Beta	STAT	P-Value	EMP1	EMP2	Pain Intensity
NRF3	rs4722585	0.214	2.525	0.013*	0.013*	0.238	↑
TOP1mt	rs7387720	-0.166	-1.942	0.054	0.054	0.683	↓

P-value (Significant) : *

Table 3.21. Linear regression analysis of genotypes in association with pain intensity.

Gene	SNP	Genotype Model	BETA	P-value	Bonferroni P-value	EMP1	EMP2	Pain Intensity
PINK1	rs10799655	Genotypic	NA	0.029*	0.895	0.701	1	-
PINK1	rs10799655	Dominant	0.213	0.015*	0.658	0.022*	0.296	↑
NFE2L3	rs4722585	Genotypic	NA	0.042*	1	0.031*	0.521	-
NFE2L3	rs4722585	Dominant	0.201	0.019*	0.506	0.013*	0.346	↑
TOP1mt	rs7387720	Genotypic	NA	0.085	1	0.399	1	-
TOP1mt	rs7387720	Dominant	-0.188	0.028*	0.740	0.032*	0.473	↓

P-value (Significant) : *

In summary, **Table 3.22** below shows all results for association with pain intensity.

Table 3.22. Summary results of Pain intensity.

				Univariate analyses			Multivariate analyses			
Gene	**SNP**	**Model**	**Associated with**	**P-value**	**pEMP1**	**Beta**	**P-value**	**pEMP1**	**pEMP2**	**Beta**
PINK1	rs10799655	Genotypic	pain intensity	0,059	0,727	NA	0.029*	0.701	1	NA
PINK1	rs10799655	Dominant	pain intensity	0.029*	0.031*	0.188	0.015*	0.022*	0.296	0.213
NRF3	rs4722585	Allelic	pain intensity	0.014*	0.014*	0.209	0.012*	0.013*	0.238	0.214.
NRF3	rs4722585	Genotypic	pain intensity	0,045*	0,059	NA	0.042*	0.031*	0.521	NA
NRF3	rs4722585	Dominant	pain intensity	0.018*	0.023	0.202	0.019*	0.013*	0.346	0.201
TOP1mt	rs7387720	Allelic	pain intensity	0.053	0.05	-0.166	0.054	0.054	0.683	-0.166
TOP1mt	rs7387720	Dominant	pain intensity	0.030*	0.033*	-0.186	0.028*	0.032*	0.473	-0.188

P-value (Significant) : *

3.10 Associations between alleles, genotype, haplotypes and mtDNA copy number

3.10.1 Univariate analysis of associations between genetic variants and mtDNA copy number

When we compared mtDNA copy number to SNP alleles, we found that there was no significant association with all variants (**Table 3.23**). SNPs in blue on the Table below had the lowest P-values for association with mtDNA copy number in the cohort but all were above significance level of 0.05. When computation for multiple correction was done, all SNPs remained insignificant for association with mtDNA copy number.

When we compared genotypes with mtDNA copy number, we found that none of the SNPs were significantly associated with mtDNA copy number. All SNPs were above significance level of $P < 0.05$.

When we compared mtDNA copy number to haplotypes (**Table 3.24**), we found that three haplotypes on chromosome ten (rs11006126| rs2306604| rs1049432 TGT in *TFAM*, rs11006132 |rs17113613| rs3740485 GGC across *TFAM* and C10orf2, and rs17113613| rs3740485| rs3740486 GCT in *C10orf2*) were significantly associated with mtDNA copy number in this cohort with $P = 0.007$, 0.001 and 0.038 respectively. They were all associated with increased mtDNA copy number as seen by the strongly positive Beta value. There was no single SNP in common between these three haplotypes, but they were all in the *TFAM and C10orf2* genes on chromosome 1. In addition, another four haplotypes in the *TFAM-C10orf2* region also showed weak associations with mtDNA copy number ($P>0.05$ but < 0.10). None of the SNPs in these haplotypes were independently associated with SN risk or mtDNA copy number in allelic or genotypic models. The *TFAM* haplotype here associated with mtDNA copy number (rs11006126| rs2306604| rs1049432 TGT) was different to the *TFAM* haplotype associated with SN (rs11006126|rs2306604|rs1049432 CGG), but had the rs2306604 G in common.

Table 3.23. Basic association test of alleles with mtDNA copy number.

Gene	SNP ID	Beta	T-test P-value	Bonferroni P-value	EMP1	EMP2
PINK1	rs10799655	-0.032	0.646	1	0.643	1
	rs650616	-0.020	0.771	1	0.770	1
	rs622525	-0.070	0.314	1	0.313	1
CCDC19	rs2501325	0.068	0.331	1	0.332	1
ZFN648	rs7554182	-0.090	0.197	1	0.197	0.988
NRF2	rs1962142	0.083	0.236	1	0.234	0.996
	rs35652124	0.120	0.085**	1	0.085	0.834
	rs2886162	-0.082	0.247	1	0.249	0.997
NFE2L3	rs4722585	-0.062	0.375	1	0.378	1
TOP1mt	rs7387720	0.005	0.942	1	0.942	1
	rs2293925	-0.002	0.982	1	0.982	1
	rs724037	-0.005	0.947	1	0.947	1
	rs11544484	-0.044	0.534	1	0.537	1
	rs2450772	0.029	0.684	1	0.688	1
TFAM	rs11006126	0.009	0.901	1	0.903	1
	rs2306604	-0.052	0.459	1	0.461	1
	rs1049432	0.088	0.207	1	0.209	0.991
	rs11006132	0.122	0.082**	1	0.083	0.822
C10orf2	rs17113613	-0.015	0.827	1	0.829	1
	rs3740485	0.119	0.091**	1	0.093	0.853
	rs3740486	0.120	0.087**	1	0.088	0.840
	rs11190787	0.132	0.062**	1	0.063	0.730
POLG	rs2856268	-0.109	0.121	1	0.120	0.926
LITAF	rs13333308	-0.123	0.077**	1	0.075	0.804
POLG2	rs9905016	-0.085	0.228	1	0.228	0.995
	rs2075551	-0.078	0.265	1	0.264	0.998

P-value (Significant) : *
P-value (Border-line Significant) : **

Table 3.24. Haplotype association with mtDNA copy number.

LOCUS	HAPLOTYPE	Beta	R2	STAT	P	pEMP1	pEMP2	SNPS	GENES
WIN1	TGA	-17.27	0.0032	-0.81	0.418	0.414	1	rs10799655|rs650616|rs622525	*PINK1*
WIN1	CAT	-5.33	0.0005	-0.33	0.745	0.754	1	rs10799655|rs650616|rs622525	*PINK1*
WIN1	TGT	6.994	0.0004	0.29	0.771	0.795	1	rs10799655|rs650616|rs622525	*PINK1*
WIN1	CGT	17.46	0.0047	0.99	0.325	0.330	1	rs10799655|rs650616|rs622525	*PINK1*
WIN2	GAC	89.9	0.0044	0.95	0.344	0.301	1	rs650616|rs622525|rs2501325	*PINK1 – CCDC19*
WIN2	ATC	1.683	8.3e-006	0.04	0.967	0.973	1	rs650616|rs622525|rs2501325	*PINK1 – CCDC19*
WIN2	GTC	29.74	0.005	1.02	0.311	0.301	1	rs650616|rs622525|rs2501325	*PINK1 – CCDC19*
WIN2	GAA	-27.63	0.008	-1.27	0.207	0.202	1	rs650616|rs622525|rs2501325	*PINK1 – CCDC19*
WIN2	ATA	-6.234	0.0006	-0.35	0.724	0.731	1	rs650616|rs622525|rs2501325	*PINK1 – CCDC19*
WIN2	GTA	10.38	0.002	0.58	0.563	0.572	1	rs650616|rs622525|rs2501325	*PINK1 – CCDC19*
WIN3	ACT	77.11	0.004	0.88	0.382	0.349	1	rs622525|rs2501325|rs7554182	*PINK1 – CCDC19-ZFN*
WIN3	TCT	-5.952	4.1e-005	-0.09	0.927	0.931	1	rs622525|rs2501325|rs7554182	*PINK1 – CCDC19-ZFN*
WIN3	AAT	-57.75	0.005	-0.96	0.336	0.335	1	rs622525|rs2501325|rs7554182	*PINK1 – CCDC19-ZFN*
WIN3	TAT	-30.19	0.009	-1.37	0.171	0.167	1	rs622525|rs2501325|rs7554182	*PINK1 – CCDC19-ZFN*
WIN3	TCC	26.53	0.005	0.99	0.322	0.325	1	rs622525|rs2501325|rs7554182	*PINK1 – CCDC19-ZFN*
WIN3	AAC	-26.47	0.005	-1.05	0.294	0.296	1	rs622525|rs2501325|rs7554182	*PINK1 – CCDC19-ZFN*
WIN3	TAC	19.33	0.007	1.20	0.232	0.214	1	rs622525|rs2501325|rs7554182	*NRF2*
WIN4	ATA	25.65	0.002	0.62	0.536	0.569	1	rs1962142|rs35652124|rs2886162	*NRF2*
WIN4	GTA	-25.28	0.011	-1.52	0.130	0.134	0.996	rs1962142|rs35652124|rs2886162	*NRF2*
WIN4	GCG	32.74	0.012	1.60	0.111	0.126	0.993	rs1962142|rs35652124|rs2886162	*NRF2*
WIN4	GTG	-5.826	0.001	-0.35	0.728	0.731	1	rs1962142|rs35652124|rs2886162	*TOP1mt*
WIN6	GAG	14.77	0.0002	0.18	0.855	0.844	1	rs7387720|rs2293925|rs724037	*TOP1mt*
WIN6	GGG	0.9037	1.1e-005	0.05	0.963	0.964	1	rs7387720|rs2293925|rs724037	*TOP1mt*

LOCUS	HAPLOTYPE	Beta	R2	STAT	P	pEMP1	pEMP2	SNPS	GENES
WIN6	AGG	-54.47	0.0043	-0.94	0.349	0.369	1	rs7387720|rs2293925|rs724037	*TOP1mt*
WIN6	GGT	10.28	9.9e-005	0.14	0.887	0.868	1	rs7387720|rs2293925|rs724037	*TOP1mt*
WIN6	AGT	3.75	0.0002	0.21	0.832	0.832	1	rs7387720|rs2293925|rs724037	*TOP1mt*
WIN7	AGC	-1.689	2.6e-006	-0.02	0.982	0.977	1	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN7	GGC	-2.386	7.9e-005	-0.13	0.899	0.906	1	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN7	GTC	-12.03	0.002	-0.64	0.521	0.529	1	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN7	GTT	10.04	0.002	0.65	0.520	0.508	1	rs2293925|rs724037|rs11544484	*TOP1mt*
WIN8	GCA	-1.012	9.61e-006	-0.04	0.965	0.960	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TCA	81.97	0.005	1.03	0.305	0.287	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TTA	5.905	0.0004	0.28	0.778	0.776	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	GCG	-5.906	0.0002	-0.18	0.857	0.851	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TCG	-17.51	0.004	-0.90	0.370	0.399	1	rs724037|rs11544484|rs2450772	*TOP1mt*
WIN8	TTG	8.299	0.001	0.47	0.636	0.670	1	rs724037|rs11544484|rs2450772	*TFAM*
WIN9	CGT	10.82	0.002	0.58	0.564	0.577	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	TGT	197	0.036	2.75	0.007*	0.014*	0.345	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	CAG	-42.99	0.007	-1.17	0.245	0.228	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	TAG	-3.122	8.6e-005	-0.13	0.895	0.889	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	CGG	12.02	0.0005	0.31	0.754	0.769	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN9	TGG	-9.996	0.002	-0.62	0.533	0.531	1	rs11006126|rs2306604|rs1049432	*TFAM*
WIN10	GTG	23.51	0.008	1.27	0.207	0.219	1	rs2306604|rs1049432|rs11006132	*TFAM*
WIN10	AGA	-14.77	0.003	-0.74	0.461	0.483	1	rs2306604|rs1049432|rs11006132	*TFAM*
WIN10	GGA	-14.9	0.00	-0.93	0.354	0.366	1	rs2306604|rs1049432|rs11006132	*TFAM-C10orf2*
WIN11	TGA	-27.87	0.0009	-0.44	0.662	0.660	1	rs1049432|rs11006132|rs17113613	*TFAM-C10orf2*
WIN11	GAA	-1.374	1.0e-005	-0.05	0.963	0.961	1	rs1049432|rs11006132|rs17113613	*TFAM-C10orf2*
WIN11	TGG	29.07	0.011	1.48	0.140	0.152	0.998	rs1049432|rs11006132|rs17113613	*TFAM-C10orf2*

LOCUS	HAPLOTYPE	Beta	R2	STAT	P	pEMP1	pEMP2	SNPS	GENES
WIN11	GAG	-26.23	0.012	-1.57	0.118	0.128	0.994	rs1049432|rs11006132|rs17113613	*TFAM-C10orf2*
WIN12	GAC	-31.42	0.001	-0.50	0.618	0.625	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	AAC	-1.418	1.2e-005	-0.05	0.961	0.963	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	GGC	116	0.054	3.42	0.001*	0.004*	0.098	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	AGC	11.52	0.001	0.53	0.598	0.613	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	GGT	4.211	0.0002	0.18	0.860	0.875	1	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN12	AGT	-31.76	0.018	-1.95	0.053**	0.053**	0.908	rs11006132|rs17113613|rs3740485	*TFAM-C10orf2*
WIN13	ACT	-5.575	0.0003	-0.23	0.816	0.818	1	rs17113613|rs3740485|rs3740486	*C10orf2*
WIN13	GCT	37.45	0.021	2.09	0.038*	0.051**	0.826	rs17113613|rs3740485|rs3740486	*C10orf2*
WIN13	GTC	-28.01	0.015	-1.74	0.084**	0.105	0.976	rs17113613|rs3740485|rs3740486	*C10orf2*
WIN14	CTG	30.03	0.017	1.85	0.066**	0.074**	0.946	rs3740485|rs3740486|rs11190787	*C10orf2*
WIN14	TCT	-28.01	0.015	-1.74	0.084**	0.105	0.976	rs3740485|rs3740486|rs11190787	*C10orf2*
WIN17	TC	-24.09	0.007	-1.21	0.228	0.233	1	rs9905016|rs2075551	*POLG2*
WIN17	CC	8.208	0.0002	0.20	0.845	0.853	1	rs9905016|rs2075551	*POLG2*
WIN17	CG	19.96	0.005	1.06	0.292	0.305	1	rs9905016|rs2075551	*POLG2*

P-value (Significant) : *
P-value (Border-line Significant) :**

3.10.2 Multivariate analysis of genetic variants vs mtDNA copy number

No multivariate analysis was conducted on mtDNA copy number since no demographic data was significantly associated with mtDNA copy number.

In summary, **Table 3.25** below shows all results for association with mtDNA copy number.

Table 3.25: Summary of genetic variants vs mtDNA copy number.

Gene	SNP	Model	Associated with	Univariate analyses			Beta
				P-value	pEMP1	pEMP2	
TFAM	rs11006126|rs2306604|rs1049432	Haplotype (TGT)	mtDNA copy number	0.007*	0.015*	0.345	197
TFAM-c10orf2 (chr10)	rs11006132|rs17113613|rs3740485	Haplotype (GGC)	mtDNA copy number	0.001*	0.003*	0.098	116
TFAM-c10orf2 (chr10)	rs11006132|rs17113613|rs3740485	Haplotype (AGT)	mtDNA copy number	0.053	0.054	0.908	-31.76
C10orf2	rs17113613|rs3740485|rs3740486	Haplotype (GCT)	mtDNA copy number	0.038*	0.051	0.826	37.45

P-value (Significant) : *

3.11 Genetic risk scores (GRS)

GRS was performed using the SNPs that were found to be significantly associated with SN, pain intensity and mtDNA copy number. GraphPad Prism v7.0 was used for statistical analysis of both categorical and continuous data.

3.11.1 GRS and SN

The risk alleles of SNPs found to be associated with SN during univariate analysis were used to compute GRS. Only rs1962142, rs724037 and rs2856268 SNPs from *NRF2*, *TOP1mt* and *POLG1* genes, respectively, were found to be significantly associated with SN. Majority of the patients in our study cohort did not carry the risk alleles of both SNPs since they had a GRS of 0. **Figure 3.9** below shows the association between GRS from 3 SNPs and SN. A P-value of 0.583 was obtained. Therefore, these three SNPs do not cumulatively contribute to SN.

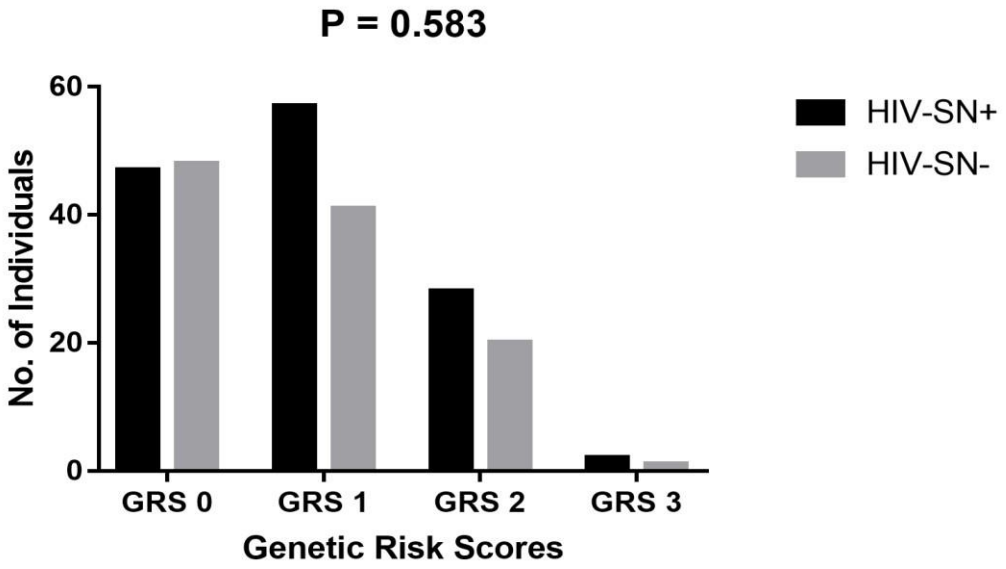


Figure 3.9. GRS vs SN. GRS 0 = absence of risk allele, GRS 1 = presence of one risk allele, GRS 2 = presence of two risk alleles, GRS 3 = presence of three risk alleles

3.11.2 GRS and pain intensity

SNPs rs4722585 (*NRF2*) and rs7387720 (*TOP1mt*) were significantly associated with pain intensity during univariate analysis. The rs10799655 SNP in *PINK1* gene was borderline significant after correction for pEMP1. Risk alleles from all three SNPs were used for calculation of GRS. With the aid of GraphPad prism v7.0, an unpaired and non-parametric T-test (Mann-Whitney test), a P-value of 0.940 (**Figure 3.10**) was computed and it was statistically not significant for association with pain intensity. Therefore, these three SNPs do not cumulatively contribute to pain intensity in SN.

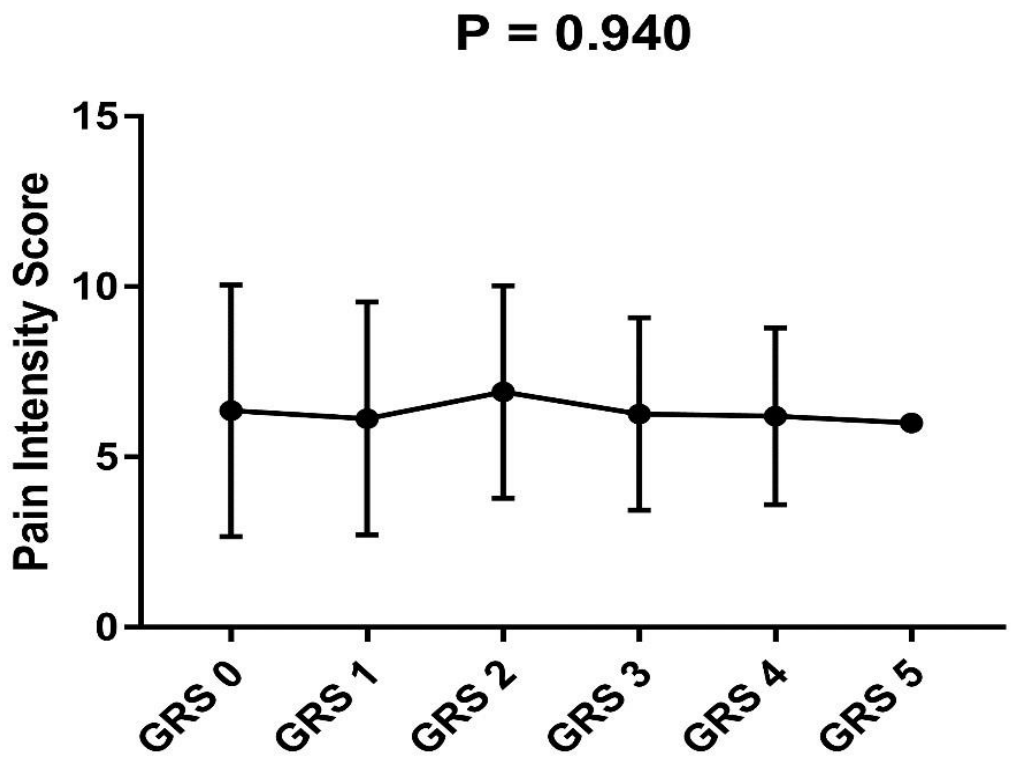


Figure 3.10: GRS vs pain intensity in patients with SN. GRS 0 = absence of risk allele, GRS 1 = presence of one risk allele, GRS 2 = presence of two risk alleles, GRS 3 = presence of three risk alleles, GRS 4 = presence of four risk alleles and GRS 5 = presence of five risk alleles.

3.11.3 GRS and mtDNA copy number

Since no SNP was significantly associated with mtDNA copy number during univariate analysis, the risk alleles of SNPs which were found to be borderline significant (P > 0.05 but < 0.1) were used for the calculation of the GRS.

Three SNPs (rs3740485, rs3740486 and rs11190787) from the *C10orf2* gene, one SNP (rs11006132) from the *TFAM* gene and one SNP (rs35652124) from the *NRF2* gene were all borderline significant for association with mtDNA copy number in HIV-positive patients exposed to d4T. Once GRS computation was complete, one-way ANOVA was employed for computation of P-value. A non-significant P-value was generated (P = 0.378) as depicted by **Figure 3.11**. Therefore these four SNPs do not cumulatively contribute to mtDNA copy number variation.

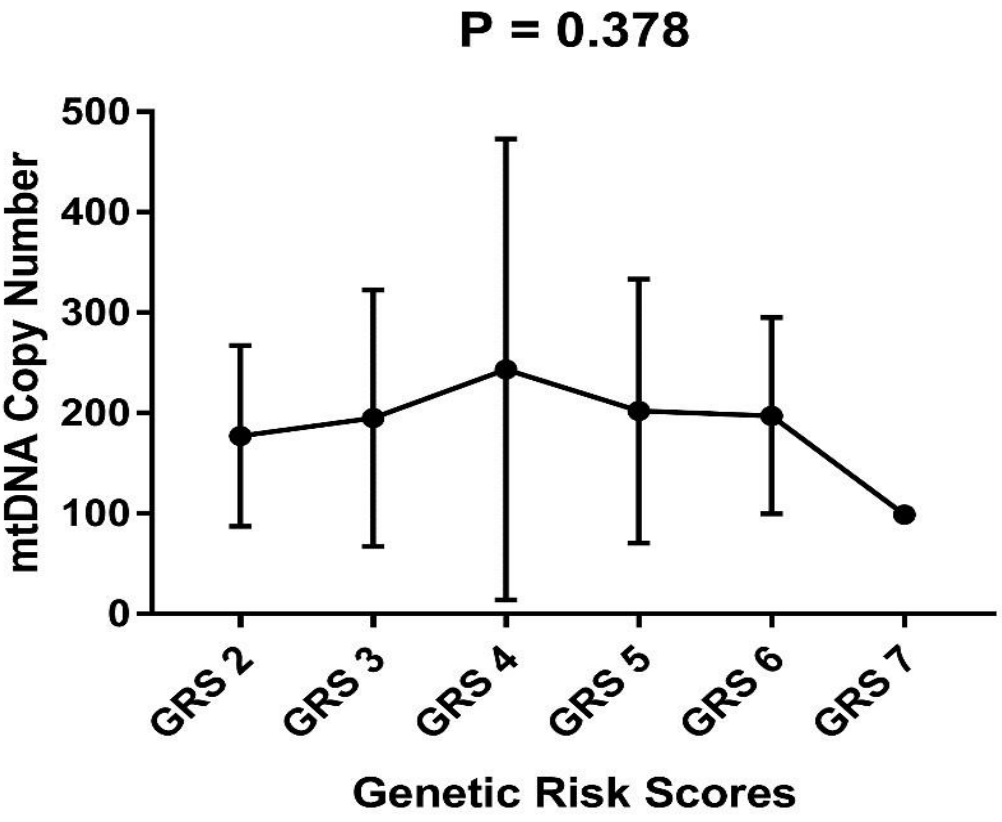


Figure 3.11: GRS vs mtDNA copy number in HIV-positive patients exposed to d4T. GRS 2 = presence of two risk alleles, GRS 3 = presence of three risk alleles, GRS 4 = presence of four risk alleles**,** GRS 5 = presence of five risk alleles, GRS 6 = presence of six risk alleles and GRS 7 = presence of seven risk alleles.

CHAPTER 4: DISCUSSION AND CONCLUSION

SN is a common neurological side effect related to use of d4T-containing ART. Signs and symptoms include a sense of pins-and-needles, loss/reduced reflexes of ankles and sense of vibration. SN is thought to arise via ART inhibition of mtDNA polymerase gamma, with subsequent effects on mtDNA copy number and mitochondrial function. The fact that not everyone using d4T experiences SN suggests a possible genetic predisposition underlying SN occurrence.

Genetic variation in *POLG* (Chiappini *et al*. 2009) and the TNFa region (Cherry *et al.* 2008; Wadley *et al*. 2014) has previously been shown to associate with SN after ART use. However, it has also been shown that genetic variation in the *POLG* gene is not an underlying cause of SN after ART use in African populations (van Oosterhout *et al.* 2013; Ojwach *et al*., 2016). The role of variation in other genes influencing mtDNA replication in SN development after ART has not been explored much, although variation in these genes were not found to be associated with SN in the GWAS performed by Leger *et al.* (2014). Variation in most of these genes has not previously been described in Black South Africans. The first aim of this study was therefore to determine whether genetic variation in candidate nuclear genes influencing mitochondrial replication was associated with the development of SN in Black South African HIV-positive individuals who have used d4T containing antiretroviral therapy.

There has been conflicting information in the literature regarding the occurrence of decreased mtDNA copy number in individuals experiencing SN after ART use. A decrease in mtDNA copy number has been demonstrated in nerve cells in patients with SN after ART (Dalakas et al. 2011), in whole blood samples from Malawian patients with SN on d4T ART (Kampira *et al.* 2014). and in association with other phenotypes of mitochondrial toxicity after ART use, including lipodystrophy (Chiappini et al. 2009) and lactic acidosis (Cote et al. 2002; Kampira et al. 2014). However Cherry et al. (2002) showed that ART-induced lactatemia was associated with decreased mtDNA copy number in adipose tissue but not in PBMC. The second aim of this study was therefore to check if mtDNA copy number was decreased in whole blood from South African individuals who experienced SN after d4T use. Thirdly, the associations between SNPs in the

genes governing mtDNA replication and variation in mtDNA copy number were examined.

We examined associations between 28 SNPs from 11 genes (*POLG, POLG2, NRF2, TFAM, C10orf2, TOP1mt, PINK1, NRF3, LITAF, CCDC19* and *ZNF648)* genes and three different phenotypes: SN, pain intensity and mtDNA copy number. These SNPs were chosen based on their reported or predicted functional effect, effect on mtDNA copy number, and/or association with SN. It was also taken into consideration whether the SNPs had MAF greater or equals to 5% in LWK, if SNP assay could be designed successfully in Sequenom software, and whether they were tag SNPs. Four of the SNPs were chosen due to their very low (although not genome-wide) significance levels in Africans with SN in the GWAS performed by Leger *et al.* (2014). The associations between these SNPs and 3 phenotypes (SN, mtDNA copy number and pain intensity), were assessed in a black South African cohort on d4T-ART for at least six months.

4.1 Genetic variation in nuclear genes of the mtDNA replication pathway in the black SA population

Our reports of MAF for these SNPs in the black South African populations is novel. The MAF for 25 of the SNPs was >5% in the cohort of Black South Africans. Our findings showed that all 26 SNPs were in Hardy-Weinberg equilibrium (HWE) and there was a slight variation in frequencies of the alleles when compared to the LWK and YRI African populations.

Our data were used to construct haplotypes in six genes: *POLG2, NRF2, C10orf2, TOP1mt, TFAM, PINK1*. In addition, haplotypes across three genes on chromosome 1 (*PINK, ZNF648* and *CCDC19*) and across two genes on chromosome 10 (*TFAM* and *c10orf2*) were constructed. Shorter haplotypes were expected in African populations as compared to non-African populations. Variation in allele frequencies and LD estimates in different populations may be due to micro-evolutionary factors, such as genetic drift, migration, population structure and natural selection.

4.2 Association of demographic factors with SN and with mtDNA copy number

Demographic variables of age and height were found to be major risk factors for SN in HIV-positive patients treated with d4T-containing regimens This was previously reported in this cohort by Wadley *et al.,* (2011) and Hendry (2015). and these associations were similar to results found by Cherry *et al.,* (2009). Interestingly length of time on d4T was not significantly associated with SN in this cohort, which could indicate that d4T mediates toxicity within a short time period. Menezes *et al.,* (2013) reported that patients who were exposed to the standard and low dose d4T had a decrease in mean mtDNA copy number/cell of 29 % and 32 %, respectively when compared with TDF after only four weeks. Kampira (2013) found that body mass index (BMI) and age (> 40) were significantly associated with d4T associated adverse effects and the possible mechanism was thought to be via increased ROS due to the synergistic effect of both NRTIs and age.

4.3 Genetic associations with SN after d4T use

We found three SNPs and two haplotypes that were significantly associated with the occurrence of SN in our cohort using univariate analysis. These included rs1962142 in *NRF2,* rs724037 in *TOP1mt* and rs2856268 in *POLG* and haplotypes rs7387720|rs2293925|rs724037 GGT in *TOP1mt* and *TFAM* rs11006126|rs2306604|rs1049432 CGG. All of these represent novel loci for associations with SN. Of these, (rs7387720|rs2293925|rs724037 GGT in *TOP1mt* and TFAM rs11006126|rs2306604|rs1049432 CGG) remained associated with SN in multivariate analysis using age and height as covariates, and may therefore be considered as potential independent susceptibility factors for SN occurrence after d4T use.

However, none of these genetic variants remained significant after corrections for multiple testing. These results are concordant with the data from Leger *et al.* (2014) who showed that susceptibility to d4T/ddI-associated neuropathy was not explained by a single genetic variant with a marked effect. As seen in GRS, the combination of the individual risk alleles of SNPs rs1962142, rs723047 and rs2856268 in genes *NRF2, TOP1mt* and *POLG,* respectively did not additively contribute to SN occurrence.

No significant associations were found between individual SNPs in *POLG2, C10orf2, TFAM* or *PINK1,* and SN after d4T use. The four SNPs replicated in this cohort from GWAS by Leger were not significantly associated with SN in this cohort. SN replication studies are important, since they can confirm whether certain genetic variants are really associated/not associated with certain clinical phenotypes, more importantly in cohorts of different ethnicity.

The rs1962142 SNP in the *NRF2* gene was protective against SN in our study in univariate analysis, and a dominant model of genotypic effect was supported. This association was not significant when covariates of age and height were taken into account using logistic regression or when corrections were made for multiple

testing. These results are contradictory to results in the literature that this SNP was associated with occurrence of SN in diabetes patients (Xu *et al.,* 2015; Wang *et al.*, 2015). Elsewhere in Finnish and Chinese (Han) populations *NRF2* SNP rs1962142 has been associated with low synthesis of NRF2 protein in breast cancer (Hartikainen *et al.,* 2012). Low NRF2 expression could potentially cause decreased antioxidants and therefore increased mitochondrial damage which may relate to mitochondrial toxicity seen in disorders such as SN and breast cancer. This mechanism could perhaps underlie any association between rs1962142 SNP and neuropathy, but would not explain associations between *NRF2* rs1962142 and lack of neuropathy seen in this study. In addition, the effects of the SNP on NRF2 expression in non-cancer patients has not been established.

The rs2856268 SNP in the *POLG* gene was associated with the development of SN in this study in univariate analysis and in a dominant model of genotypic effect. This association was not significant when covariates of age and height were taken into account using logistic regression or when corrections were made for multiple testing. This adds to suggestions by Van Oosterhout *et al.* (2013) and Ojwach *et al.* (2016) that SNPs in *POLG* may not be common pathogenic determinants of d4T-associated mitochondrial toxicity in African populations. The R964C and E1143G *POLG* gene SNPs were found to be absent in Zulu speaking HIV-positive patients exposed to d4T containing ART (Ojwach *et al.*, 2016). Ojwach *et al*., (2016) also found that the *POLG* SNPs associated with Hyperlactatemia, Lactic Acidosis, Lipodystrophy, Peripheral Neuropathy in American, Europeans and Asians were absent in their Zulu cohort. In Van Oosterhout *et al.* (2013) neither the R964C nor E1143G SNPs of *POLG* were present among HIV-positive patients from Malawi exposed to d4T containing ART. In addition, when Van Oosterhout *et al.* (2013) sequenced the entire *POLG* gene in ten Malawian patients with severe d4T side effects, they found no mutations.

The lack of association in the current study between rs2856268 in *POLG* and SN in multivariate analysis, is different to observations that this SNP was associated

with the development of diabetic polyneuropathy in Russian patients diagnosed with type 1 diabetes mellitus (Spitsina *et al.* 2009; Malyarchuk *et al.*, 2011). The rs2856268 SNP is located in the 5'- promoter region of the *POLG* gene. The patients with diabetic neuropathy harbouring the rs2856268 SNP C allele had a decreased synthesis of the *POLG* enzyme as compared to the control group (Spitsina *et al.*, 2009). In the current study the SNP was not associated with decreased mtDNA copy number; we did not measure direct effect of the SNP on *POLG* expression or activity.

One SNP rs724037 (allele G) within the *TOP1mt* gene was found to be significantly associated with the development of SN in our cohort but only in univariate analysis in an allelic model. This association was not significant when covariates of age and height were taken into account using logistic regression or when corrections were made for multiple testing. This SNP was therefore not considered as independent risk factor for SN occurrence after d4T use. No associations between rs724037 (allele G) and any type of neuropathy has been previously reported.

The *TOP1mt* haplotype (GGT rs7387720|rs2293925|rs724037) was significantly associated with protection against SN in our cohort in both univariate and multivariate analysis. This protective haplotype included rs724037 allele T, which is consistent with the association of rs724037 (risk allele G) with SN discussed above. The rs7387720 SNP within the protective *TOP1mt* haplotype, has previously been associated with SN in Diabetes mellitus type 1 patients from Russia (Swan *et al.*, 2015), consistent with association with SN in this study. This SNP rs7387720 (allele G), was also significantly associated with pain intensity in both univariate and multivariate analysis in this study. Both rs724037 (allele G SNP) and rs7387720 have previously been associated with diabetic kidney disease and end-stage renal disease in Caucasian Europeans (Swan *et al.*, 2015). How the rs724037 SNP, an intron variant, causes dysfunction of the topoisomerase enzyme is not known. Wang *et al.* (2011) found that the minor allele of the *TOP1mt* rs2293925 SNP was associated with TOP1mt protein truncation in patients diagnosed with mitochondrial disorders.

Lastly, the CGG haplotype from the *TFAM* (Mitochondrial transcription factor A) gene was associated with the development of SN in the current study in univariate and multivariate analyses. The haplotype was constructed with alleles from the following *TFAM* gene SNPs, rs11006126 C, rs2306604 G and rs1049432 G. The rs11006126 SNP was not individually associated with SN development in this cohort, but was present in TFAM haplotypes significantly associated with mtDNA copy number in this cohort (discussed later). The minor allele C was also associated with decreased mtDNA copy number amount in Chinese patients diagnosed with depression (Cai *et al.,* 2015), but has not previously been associated with any neuropathy. The single SNP rs2306604 (allele A) located in intron 4 of the *TFAM* gene was associated with the risk of developing Parkinson's Disease (PD) in French patients (Bertram *et al.,* 2007; Laumet *et al.,* 2010) and also in British and Scandinavian cohorts (Gatt *et al.,* 2013). Belin *et al.,* (2007) showed that rs2306604 A was associated with an increased risk of PD and Alzheimer's disease (AD) development, using case control data from Italy, Germany and Switzerland. Similarly, rs1049432 (allele T) was associated with PD and AD (Taherzadeh-Fard *et al.,* 2011).

None of the SNPs and haplotypes described above confer their protective or susceptibility effects against SN development after d4T use, by changes in mtDNA copy number, as none of them were significantly associated with mtDNA copy number in this cohort.

4.4 Genetic associations with pain intensity after d4T use

Individual SNPs in *NRF3, TOP1mt* and *PINK1* genes were significantly associated with pain intensity in allelic and/or genotypic models in univariate analyses. The NRF3 SNP remained significantly associated with pain intensity in multivariate analysis. However, none of these genetic variants remained significant after corrections for multiple testing. Risk alleles from rs4722585 (*NRF3*), rs7387720 (*TOP1mt*) and rs10799655 (*PINK1*) were used for the computation of GRS. There was a no significant association of the scores with pain intensity, suggesting no additive effect for these risk alleles.

The rs4722585 SNP in *NRF3* gene was significantly associated with an increase in pain intensity in our cohort of HIV-positive patients diagnosed with SN and exposed to d4T. In a GWAS by Leger *et al.,* (2014) this SNP was found to be non-significant for association with SN at genome-wide level, but had a very low P-value ($P = 4.4 \times 10^{-06}$) in Black individuals, in association with grade 2 SN, where grade 2 was defined as sensory alteration or paraesthesia causing greater than minimal interference with usual social and functional activities. How the *NFR3* SNP exerts its effect is currently unknown since it is a variant located in the regulatory region of the gene and not much literature on this variant is currently available. *NRF3* gene like *NRF2* encodes a transcription factor that regulates the synthesis of antioxidants which maintains the normal physiological function of the mitochondrial. Any deleterious impairment of this pathway more specifically in the neurons can induce pain via starvation of the nerve cells due to low ATP production as a result of oxidants damaging the mitochondria (Taherzadeh-Fard *et al.,* 2011). If the rs4722585 causes a reduction in the synthesis of NRF3 protein, this might decrease antioxidant response and increase ROS damage, therefore influencing pain occurrence and intensity.

TOP1mt SNP rs7387720 was significantly associated with decreased pain intensity in HIV-positive patients with SN in our cohort. It was also associated with SN occurrence in our cohort and this SNP was also discussed in the section above. The genetic variant was associated with end stage renal disease in patients with DM1 (Swan *et al.,* 2015). *TOP1mt* does not primarily regulate any known pain mechanism.

SNP rs10799655 in the *PINK1* gene was significantly associated with pain intensity in a dominant model in univariate analysis in our cohort. In a study by Franks *et al.* (2008) the rs10799655 was associated with an increase in PINK1 mRNA levels in Danish diabetics. PINK1 protein maintains the normal physiological function of the mitochondria, and faulty mitochondria are targeted by PINK1 for destruction.

4.5 mtDNA copy number association with SN

mtDNA copy number has been used as a biomarker for many HIV and ART-related neurological complications including SN (Côté *et al.,* 2002; Cherry *et al.,* 2002; Kampira *et al.,* 2014). 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). Kampira *et al.,* (2014) found significantly lowered mtDNA copy number in whole blood samples in association with ART- induced SN and lactatemia, but not lipodystrophy. It remains debatable whether ART -induced SN or other mitochondrial toxicities are associated with mtDNA depletion that can be observed in blood cells.

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. Diagnosis of SN in Kampira *et al.,* (2014) was based on characteristic symptoms that had started after the initiation of ART, while in our study some patients already had SN before ART. Cherry *et al*., (2002) showed that ART-induced lactatemia was associated with decreased mtDNA copy number in adipose tissue but not in PBMC. 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 may not be as accurate as measurement in PBMC depleted of platelets.

4.6 Genetic associations with mtDNA copy number after d4T use

Although our data did not support a decrease in mtDNA copy number in SN, we did find significant associations between genetic variants and mtDNA copy number in the total cohort. There were no individual alleles or genotypes associated with mtDNA copy number in the cohort. Combinations of alleles in GRS also did not yield any significant association with mtDNA copy number.

Three haplotypes in the TFAM-c10orf2 region on chromosome ten were significantly associated with increased mtDNA copy number including TFAM haplotype TGT (rs11006126, rs2306604, rs1049432; TFAM-c10orf2 haplotype GGC (rs11006132, rs17113613, rs3740485); and C10orf 2 haplotype GCT (rs17113613, rs3740485, rs3740486). Since no demographic data were associated with mtDNA copy number, no multivariate analyses were performed with these haplotypes.

Two of these haplotypes had *C10orf2* rs17113613 (allele G) and *C10orf2* rs3740485 (allele C) in common, strengthening the hypothesis that these variants relate to increased mtDNA copy number. However, these SNPs were not individually associated with mtDNA copy number, suggesting that there may be a functional underlying SNP in LD with these haplotypes that was not assessed in this study.

One of our haplotypes consisted of SNPs only in the *TFAM* gene. This rs11006126| rs2306604| rs1049432 TGT in *TFAM* was significantly associated with an increase in mtDNA copy number in our cohort. Similarly, the TFAM rs11006126 locus has also been highly significantly associated with mtDNA copy number in Chinese and UK cohorts (Cai *et al.,* 2015). The rs2306604 risk (allele A) has been associated with mtDNA depletion in neurodegenerative disorders such as Alzheimer's, Parkinson's and Huntington's disease (Belin *et al.,*2007; Gatt *et al.,* 2013; Taherzadeh-Fard *et al.,* 2011). This is consistent with our finding that the rs2306604 (allele G) (in haplotype) was associated with increased amounts of mtDNA. The risk (allele rs1049432 T was found to be associated with mitochondrial dysfunction and age of early onset in patients diagnosed with Huntington's disease (Taherzadeh-Fard *et al.,* 2011). Our findings of higher mtDNA copy number in association with a haplotype containing this T allele are in contrast to their findings of mitochondrial dysfunction in association with this allele. Taherzadeh-Fard *et al.* (2011) also showed that SNPs rs11006126 and rs1049432 were in high LD with each other in their German cohort, similar to the high LD observed across these SNPs in our African cohort.

The haplotype (rs17113613| rs3740485| rs3740486) GCT in *C10orf2,* and the related haplotype (rs11006132 |rs17113613| rs3740485) GGC across *TFAM* and *C10orf2,* were associated with an elevation in mtDNA in our cohort samples. The rs17113613 A has previously been reported in a patient with symptoms of mtDNA stability defects (Naimi *et al.,* 2006). This is consistent with our association of the G allele with increased mtDNA. However, the risk alleles of rs3740485 C and rs3740486 T SNPs in *C10orf2* were associated with Progressive External Opthalmoplegia (PEO) in an Indian cohort (Singh *et al.,* 2012). These results are in contrast with our results which show association with increased mtDNA. In addition, *TFAM* rs11006132-G was associated with Huntington's disease risk Taherzadeh-Fard *et al.,* 2011) which may indicate association with lower mtDNA copy number, again in contrast with our results which show association with increased mtDNA .

Whilst possession of these three haplotypes was associated with increased mtDNA copy number in our cohort, they were not related to susceptibility to or protection from ART-induced SN in this cohort. These haplotypes could potentially play a role in protection against other ARV/NRTI mitochondrial toxicities, or other diseases where mtDNA copy number is important e.g. diabetic neuropathy and chemotherapy induced neuropathy

4.7 Study limitations

The cohort size was relatively small (n = 263) for a genetic association study. Smaller cohort sizes carry with them lower statistical power for association with clinical phenotypes. Similar studies should therefore be reproduced in a larger sample.

When diagnosing SN, patients were not categorised as having SN prior or post d4T exposure, this makes it difficult to understand whether HIV or d4T induced the SN. Therefore, in future studies, SN diagnosis should occur both before and after ART initiation.

A different African population (LWK) was used as a proxy to choose SNPs of interest. It is possible that the LD in our cohort and that of the proxy population differs, which may affect our results especially when using tag SNPs for association with respective clinical phenotype. Tag SNPs showing a strong LD in LWK may not be when genotyped in South Africans. This is further proved by the fact that many genetic variants that are significantly associated with certain disorders in Caucasian populations, become non-significant when replicated in African populations (Hendry, 2015).

4.8 Future work

We focussed on 28 SNPs in *POLG1, POLG2, NRF2, C10orf2, TOP1mt, TFAM, PINK1, CCDC19, LITAF, NRF3*and *ZNF648* genes in this study. While no genetic variants remained significantly associated with SN or pain intensity after correction for multiple testing, our results suggested that variants in the *TFAM-c10orf2* region may influence mtDNA copy number in Black South Africans on ART. It would be interesting to test these results in a larger sample size of healthy Black South Africans. Further studies of these loci using increased SNP density across the genes may be required (since this will ensure that a larger region of the genes of interest is interrogated) and functional studies would also clarify SNP role. Workalemahu *et al.,* (2017) recently published results of GWAS for mtDNA copy number and candidate genes arising from their study may be interesting to consider in future studies related to mtDNA copy number in Black South Africans.

Overall our findings of several genetic variants that are significantly associated with SN and/or mtDNA copy number support other observations in the literature of roles for these loci in similar phenotypes. The role of these variants in other phenotypes of mitochondrial toxicity apart from SN, such as lactic acidosis or lipodystrophy.

However, the exact role of specific variants in terms of increased or decreased mtDNA copy number and susceptibility to SN remains in question, as several of our results had the opposite effect to those reported in the literature. It seems unlikely that the SNPs would have different effects in different disorders if the

disorders all share the common mechanism of mtDNA depletion. However, this assumption of common mechanism of mtDNA depletion may not be true, especially since several of the articles we were comparing our results to had not directly measured mtDNA copy number. Mitochondrial function or damage might be more important than mtDNA copy number. Functional mtDNA assays such as luciferase assay and lactic acid measurement could be performed since both can measure mitochondrial activity and mitochondrial dysfunction respectively.

Only 28 SNPs in eleven candidate genes were considered here and there are obviously vastly many more SNPs and genes that could have been chosen. For example, there are many more genes in the human nuclear genome that play a role in mtDNA regulation. There are approximately 1,500 genes that specify nuclear-encoded mitochondrial proteins some of which are associated with AIDS progression (Hendrickson *et al.,* 2010) and which contribute to ART-related complications. However, GWAS by Leger *et al*., (2014) failed to detect any single genetic variants that could account for susceptibility to d4T/ddI-associated neuropathy, suggesting that it may be unlikely that other mtDNA-related genes may be very strongly implicated in SN.

Sequencing of genes that regulate and/maintain the mitochondria might aid in developing SNP database representing genetics of the South African population. Inclusion of individuals of varying South African ethnicity would greatly assist in the understanding of population stratification and its role in genetic association studies. Although the expensive nature of sequencing might serve as a limitation, novel polymorphism may also be discovered. All this requires a large cohort size, which will improve the statistical power during analysis. The availability of a genetic database specifically for South Africans might help when choosing SNPs of interest, rather than using a proxy population.

Since the functional role of some of the genetic variants investigated in the current study are unknown, functional studies can also be conducted to help understand the impact of these variants. Expression studies of the mitochondrial replisome machinery proteins may also aid in understanding as to how mitochondrial toxicity due to genetic variants and d4T come about.

Even though d4T is no longer in use much in SA, this study remains relevant since NRTIs are still being used to treat disorders such as diabetic neuropathy and chemotherapy-induced neuropathy. Significant genetic variants in our study could be important for other phenotypes related to mtDNA copy number e.g. AD, PD, diabetes, athletic ability and cancer.

REFERENCES

Acharjee, S., Noorbakhsh, F., Stemkowski, P.L., Olechowski, C., Cohen, É.A., Ballanyi, K., Kerr, B., Pardo, C., Smith, P.A. and Power, C., 2010. “HIV-1 viral protein R causes peripheral nervous system injury associated with in vivo neuropathic pain”. The FASEB Journal, vol.24, no. 11, pp.4343-4353.

Affandi, J.S., Price, P., Imran, D., Yunihastuti, E., Djauzi, S. & Cherry, C.L. 2008, "Can we predict neuropathy risk before stavudine prescription in a resource-limited setting?", AIDS Research and Human Retroviruses, vol. 24, no. 10, pp. 1281-1284.

Apostolova, N., Blas-García, A. & Esplugues, J.V. 2011a, "Mitochondrial interference by anti-HIV drugs: mechanisms beyond Pol-γ inhibition", Trends in Pharmacological Sciences, vol. 32, no. 12, pp. 715-725.

Apostolova, N., Blas-Garcia, A. & V Esplugues, J. 2011b, "Mitochondrial toxicity in HAART: an overview of in vitro evidence", Current Pharmaceutical Design, vol. 17, no. 20, pp. 2130-2144.

Arnold, J.J., Sharma, S.D., Feng, J.Y., Ray, A.S., Smidansky, E.D., Kireeva, M.L., Cho, A., Perry, J., Vela, J.E. & Park, Y. 2012, "Sensitivity of mitochondrial transcription and resistance of RNA polymerase II dependent nuclear transcription to antiviral ribonucleosides", PLoS Pathog, vol. 8, no. 11, pp. e1003030.

Atema, E., van Oers, K. & Verhulst, S. 2013, "GAPDH as a control gene to estimate genome copy number in Great Tits, with cross-amplification in Blue Tits", Ardea, vol. 101, no. 1, pp. 49-54.

Austin, S. & St-Pierre, J. 2012, "PGC1alpha and mitochondrial metabolism--emerging concepts and relevance in ageing and neurodegenerative disorders", Journal of cell science, vol. 125, no. Pt 21, pp. 4963-4971.

Bailey, C.M., Kasiviswanathan, R., Copeland, W.C. & Anderson, K.S. 2009. “R964C mutation of DNA polymerase γ imparts increased stavudine toxicity by

decreasing nucleoside analog discrimination and impairing polymerase activity." Antimicrobial Agents and Chemotherapy, Vol. *53, no.* 6, pp.2610-2612.

Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. 2005, "Haploview: analysis and visualization of LD and haplotype maps", Bioinformatics (Oxford, England), vol. 21, no. 2, pp. 263-265.

Bertram, L., McQueen, M.B., Mullin, K., Blacker, D. & Tanzi, R.E. 2007, "Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database", Nature genetics, vol. 39, no. 1, pp. 17-23.

Bindu, A. & Anusha, P. 2011. "Adverse effects of highly active anti-retroviral therapy (HAART)". Journal Antiviral Antiretroviral, Vol. *3*, pp.060-064.

Birkus, G., Hitchcock, M.J. & Cihlar, T. 2002. "Assessment of mitochondrial toxicity in human cells treated with tenofovir: comparison with other nucleoside reverse transcriptase inhibitors." Antimicrobial Agents and Chemotherapy, Vol. *46, no.* 3, pp.716-723.

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, vol. 1, no. 3, pp.217-224.

Brennan, J., Kuhns, L.M., Johnson, A.K., Belzer, M., Wilson, E.C. & Garofalo, R. 2012, "Syndemic theory and HIV-related risk among young transgender women: the role of multiple, co-occurring health problems and social marginalization", American Journal of Public Health, vol. 102, no. 9, pp. 1751-1757.

Brinkman, K., Hofstede, H.J., Burger, D.M., Smeitink, J.A. & Koopmans, P.P. 1998, "Adverse effects of reverse transcriptase inhibitors: mitochondrial toxicity as common pathway", Aids, vol. 12, no. 14, pp. 1735-1744.

Brinkman, K., Smeitink, J.A., Romijn, J.A. & 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, vol. 354, no. 9184, pp. 1112-1115.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L. and Vandesompele, J., 2009. The MIQE guidelines: minimum information for publication of quantitative realtime PCR experiments. Clinical Chemistry, 55(4), pp.611-622

Cai, N., Li, Y., Chang, S., Liang, J., Lin, C., Zhang, X., Liang, L., Hu, J., Chan, W. & Kendler, K.S. 2015, "Genetic Control over mtDNA and Its Relationship to Major Depressive Disorder", Current Biology, vol. 25, no. 24, pp. 3170-3177.

Carling, P.J., Cree, L.M. & Chinnery, P.F. 2011, "The implications of mitochondrial DNA copy number regulation during embryogenesis", Mitochondrion, vol. 11, no. 5, pp. 686-692.

Chang, D.D. & Clayton, D.A. 1985. “Priming of human mitochondrial DNA replication occurs at the light-strand promoter.” *Proceedings of the National Academy of Sciences*, Vol. *82*, *no.* 2, pp.351-355.

Chen, X., Goudsmit, J. & van der Kuyl, Antoinette C 2002, "Lack of correlation between length variation in the DNA polymerase γ gene CAG repeat and lactic acidosis or neuropathy during antiretroviral treatment", AIDS Research and Human Retroviruses, vol. 18, no. 8, pp. 531-534.

Cherry, C.L., Affandi, J.S., Imran, D., Yunihastuti, E., Smyth, K., Vanar, S., Kamarulzaman, A. & Price, P. 2009, "Age and height predict neuropathy risk in patients with HIV prescribed stavudine", Neurology, vol. 73, no. 4, pp. 315-320.

Cherry, C.L., Gahan, M.E., McArthur, J.C., Lewin, S.R., Hoy, J.F. & Wesselingh, S.L. 2002, "Exposure to dideoxynucleosides is reflected in lowered mitochondrial DNA in subcutaneous fat", JAIDS- vol. 30, no. 3, pp. 271-277.

Cherry, C.L., Rosenow, A., Affandi, J.S., McArthur, J.C., Wesselingh, S.L. & 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, vol. 24, no. 2, pp. 117-123.

Cherry, C.L., Wesselingh, S.L., Lal, L. & McArthur, J.C. 2005, "Evaluation of a clinical screening tool for HIV-associated sensory neuropathies", Neurology, vol. 65, no. 11, pp. 1778-1781.

Chew, C., Cherry, C., Imran, D., Yunihastuti, E., Kamarμlzaman, A., Varna, S., Ismail, R., Phipps, M., Aghafar, Z. & Gut, I. 2011, "Tumour necrosis factor haplotypes associated with sensory neuropathy in Asian and Caucasian human immunodeficiency virus patients", Tissue Antigens, vol. 77, no. 2, pp. 126-130.

Chiappini, F., Teicher, E., Saffroy, R., Pham, P., Falissard, B., Barrier, A., Chevalier, S., Debuire, B., Vittecoq, D. & Lemoine, A. 2004. "Prospective evaluation of blood concentration of mitochondrial DNA as a marker of toxicity in 157 consecutively recruited untreated or HAART-treated HIV-positive patients." Laboratory Investigation, Vol. 84, no. 7, p.908.

Chiappini, F., Teicher, E., Saffroy, R., Debuire, B., Vittecoq, D. & Lemoine, A. 2009, "Relationship between polymerase gamma (POLG) polymorphisms and antiretroviral therapy-induced lipodystrophy in HIV-1 infected patients: a case-control study", Current HIV research, vol. 7, no. 2, pp. 244-253.

Cohen, B.H. & Naviaux, R.K. 2010, "The clinical diagnosis of POLG disease and other mitochondrial DNA depletion disorders", Methods, vol. 51, no. 4, pp. 364-373.

Cooke Bailey, J.N. and Igo, R.P. 2016. "Genetic risk scores." *Current Protocols in Human Genetics*, pp.1-29.

Cooke Bailey, J.N., Hoffman, J.D., Sardell, R.J., Scott, W.K., Pericak-Vance, M.A. & Haines, J.L. 2016, "The application of genetic risk scores in age-related macular degeneration: a review", Journal of clinical medicine, vol. 5, no. 3, pp. 31.

Cooray, H.C., Blackmore, C.G., Maskell, L. and Barrand, M.A., 2002. "Localisation of breast cancer resistance protein in micro vessel endothelium of human brain."

Copeland, W.C. 2010, "The mitochondrial DNA polymerase in health and disease" in Genome Stability and Human Diseases Springer, pp. 211-222.

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. 2002, "Changes in mitochondrial DNA as a marker of nucleoside toxicity in HIV-infected patients", New England Journal of Medicine, vol. 346, no. 11, pp. 811-820.

Cummins, J. 1998, "Mitochondrial DNA in mammalian reproduction", Reviews of reproduction, vol. 3, no. 3, pp. 172-182.

Datta, S., Ray, A., Roy, R. & Roy, B. 2016, "Association of DNA sequence variation in mitochondrial DNA polymerase with mitochondrial DNA synthesis and risk of oral cancer", Gene, vol. 575, no. 2, pp. 650-654.

Davies, M., Phiri, S., Wood, R., Wellington, M., Cox, V., Bolton-Moore, C., Timmerman, V., Moultrie, H., Ndirangu, J. & Rabie, H. 2013, "Temporal trends in the characteristics of children at antiretroviral therapy initiation in southern Africa: The IeDEA-SA Collaboration", PLoS One, vol. 8, no. 12, pp. e81037.

De Mendoza, C., De Ronde, A., Smolders, K., Blanco, F., Garcia-Benayas, T., De Baar, M., Fernández-Casas, P., González-Lahoz, J. & Soriano, V. 2004, "Changes in mitochondrial DNA copy number in blood cells from HIV-infected patients undergoing antiretroviral therapy", AIDS Research and Human Retroviruses, vol. 20, no. 3, pp. 271-273.

Dickinson, A., Yeung, K., Donoghue, J., Baker, M., Kelly, R.D., McKenzie, M., Johns, T. & John, J.S. 2013, "The regµlation of mitochondrial DNA copy number in glioblastoma cells", Cell Death & Differentiation, vol. 20, no. 12, pp. 1644-1653.

El-Hattab, A.W. & Scaglia, F. 2013, "Mitochondrial DNA depletion syndromes: review and updates of genetic basis, manifestations, and therapeutic options", Neurotherapeutics, vol. 10, no. 2, pp. 186-198.

Falkenberg, M., Larsson, N. & Gustafsson, C.M. 2007, "DNA replication and transcription in mammalian mitochondria", Annu.Rev.Biochem., vol. 76, pp. 679699.

Fish, J., Raule, N. & Attardi, G. 2004, "Discovery of a major D-loop replication origin reveals two modes of human mtDNA synthesis", Science (New York, N.Y.), vol. 306, no. 5704, pp. 2098-2101.

Franks, P.W., Scheele, C., Loos, R.J., Nielsen, A.R., Finucane, F.M., Wahlestedt, C., Pedersen, B.K., Wareham, N.J. & Timmons, J.A. 2008, "Genomic variants at the PINK1 locus are associated with transcript abundance and plasma non esterified fatty acid concentrations in European whites", FASEB journal: official publication of the Federation of American Societies for Experimental Biology, vol. 22, no. 9, pp. 3135-3145.

Gallant, J.E., Staszewski, S., Pozniak, A.L., DeJesus, E., Suleiman, J.M., Miller, M.D., Coakley, D.F., Lu, B., Toole, J.J. & Cheng, A.K. 2004, "Efficacy and safety of tenofovir DF vs stavudine in combination therapy in antiretroviral-naive patients: a 3-year randomized trial", JAMA, vol. 292, no. 2, pp. 191-201.

Gatt, A.P., Jones, E.L., Francis, P.T., Ballard, C. & Bateman, J.M. 2013, "Association of a polymorphism in mitochondrial transcription factor A (TFAM) with Parkinson's disease dementia but not dementia with Lewy bodies", Neuroscience Letters, vol. 557, pp. 177-180.

Gegg, M.E., Cooper, J.M., Schapira, A.H. & Taanman, J.W. 2009. “Silencing of PINK1 expression affects mitochondrial DNA and oxidative phosphorylation in dopaminergic cells.” *Plos One,* Vol. *4, no.* 3, p.e4756.

Gonzalez-Duarte, A., Cikurel, K. & Simpson, D.M. 2007, "Managing HIV peripheral neuropathy", Current HIV/AIDS Reports, vol. 4, no. 3, pp. 114-118.

Grady, J.P., Murphy, J.L., Blakely, E.L., Haller, R.G., Taylor, R.W., Turnbull, D.M. & Tuppen, H.A. 2014, "Accurate measurement of mitochondrial DNA deletion level and copy number differences in human skeletal muscle", PloS one, vol. 9, no. 12, pp. e114462.

Hahn, K. & Husstedt, I. 2008, "HIV Associated Neuropathies", Nerve nheilkunde, vol. 27, no. 10, pp. 889.

Hao, S. 2013, "The Molecular and Pharmacological Mechanisms of HIV-Related Neuropathic Pain", Current neuropharmacology, vol. 11, no. 5, pp. 499-512.

Hartikainen, J.M., Tengstrom, M., Kosma, V.M., Kinnµla, V.L., Mannermaa, A. & Soini, Y. 2012, "Genetic polymorphisms and protein expression of NRF2 and Sulfiredoxin predict survival outcomes in breast cancer", Cancer Research, vol. 72, no. 21, pp. 5537-5546.

Hendrickson, S.L., Lautenberger, J.A., Chinn, L.W., Malasky, M., Sezgin, E., Kingsley, L.A., Goedert, J.J., Kirk, G.D., Gomperts, E.D. & Buchbinder, S.P. 2010, "Genetic variants in nuclear-encoded mitochondrial genes influence AIDS progression", PloS One, vol. 5, no. 9, pp. e12862.

Hendry, L., Lombard, Z., Wadley, A. & 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), vol. 63, no. 1, pp. 27-30.

Hendry, L.M. 2015, Genetics of HIV-associated sensory neuropathy in black Southern Africans, MSc dissertation. Faculty of Health Sciences, University of the Witwatersrand

Hider, S.L., Thomson, W., Mack, L.F., Armstrong, D.J., Shadforth, M. & Bruce, I.N. 2008, "Polymorphisms within the adenosine receptor 2a gene are associated with adverse events in RA patients treated with MTX", Rheumatology (Oxford, England), vol. 47, no. 8, pp. 1156-1159.

Hudson, G. & Chinnery, P.F. 2006, "Mitochondrial DNA polymerase-gamma and human disease", Human Molecular Genetics, vol. 15 Spec No 2, pp. R244-52.

Hudson, G., Deschauer, M., Taylor, R.W., Hanna, M.G., Fialho, D., Schaefer, A.M., He, L.P., Blakely, E., Turnbull, D.M. & Chinnery, P.F. 2006, "POLG1, C10ORF2, and ANT1 mutations are uncommon in sporadic progressive external ophthalmoplegia with multiple mitochondrial DNA deletions", Neurology, vol. 66, no. 9, pp. 1439-1441.

Hughes, R. 2002, "Systematic reviews of treatment for inflammatory demyelinating neuropathy", Journal of Anatomy, vol. 200, no. 4, pp. 331-339.

Kamerman, P.R., Wadley, A.L. & Cherry, C.L. 2012, "HIV-associated sensory neuropathy: risk factors and genetics", Current pain and headache reports, vol. 16, no. 3, pp. 226-236.

Kampira, E. 2013, Pharmacogenetics of Stavudine: Role of genetic variation in mitochondrial DNA and Polymerase gamma among adult Malawian HIV/AIDs patients, PhD thesis. Faculty of Health Sciences, University of Cape Town

Kampira, E., Dzobo, K., Kumwenda, J., van Oosterhout, J.J., Parker, M.I. & 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, vol. 18, no. 7, pp. 438-445.

Kampira, E., Kumwenda, J., Van Oosterhout, J.J. & Dandara, C. 2013, "Mitochondrial subhaplogroups and differential risk of stavudine-induced lipodystrophy in Malawian HIV/AIDS patients", Pharmacogenomics, vol. 14, no. 16, pp. 1999-2004.

Kane, L.A., Lazarou, M., Fogel, A.I., Li, Y., Yamano, K., Sarraf, S.A., Banerjee, S. & Youle, R.J. 2014, "PINK1 phosphorylates ubiquitin to activate Parkin E3 ubiquitin ligase activity", The Journal of Cell Biology, vol. 205, no. 2, pp. 143-153.

Kazlauskaite, A., Kondapalli, C., Gourlay, R., Campbell, D.G., Ritorto, M.S., Hofmann, K., Alessi, D.R., Knebel, A., Trost, M. & Muqit, M.M. 2014, "Parkin is activated by PINK1-dependent phosphorylation of ubiquitin at Ser65", The Biochemical journal, vol. 460, no. 1, pp. 127-139.

Keswani, S.C., Pardo, C.A., Cherry, C.L., Hoke, A. & McArthur, J.C. 2002, "HIV-associated sensory neuropathies", AIDS, vol. 16, no. 16, pp. 2105-2117.

Kukat, A., Kukat, C., Brocher, J., Schäfer, I., Krohne, G., Trounce, I.A., Villani, G. & Seibel, P. 2008, "Generation of ρ0 cells utilizing a mitochondrially targeted restriction endonuclease and comparative analyses", Nucleic Acids Research, vol. 36, no. 7, pp. e44.

Laumet, G., Chouraki, V., Grenier-Boley, B., Legry, V., Heath, S., Zelenika, D., Fievet, N., Hannequin, D., Delepine, M. & Pasquier, F. 2010, "Systematic analysis of candidate genes for Alzheimer's disease in a French, genome-wide association study", Journal of Alzheimer's Disease, vol. 20, no. 4, pp. 1181-1188.

Lee, H. & Wei, Y. 2005, "Mitochondrial biogenesis and mitochondrial DNA maintenance of mammalian cells under oxidative stress", The International Journal of Biochemistry & Cell Biology, vol. 37, no. 4, pp. 822-834.

Leger, P.D., Johnson, D.H., Robbins, G.K., Shafer, R.W., Clifford, D.B., Li, J., McLaren, P.J. & Haas, D.W. 2014, "Genome-wide association study of peripheral neuropathy with D-drug-containing regimens in AIDS Clinical Trials Group protocol 384", Journal of Neurovirology, vol. 20, no. 3, pp. 304-308.

Lehmann, H.C. & Hoke, A. 2010, "Schwann cells as a therapeutic target for peripheral neuropathies", CNS & neurological disorders drug targets, vol. 9, no. 6, pp. 801-806.

Lewis, W., Kohler, J.J., Hosseini, S.H., Haase, C.P., Copeland, W.C., Bienstock, R.J., Ludaway, T., McNaught, J., Russ, R., Stuart, T. & Santoianni, R. 2006, "Antiretroviral nucleosides, deoxynucleotide carrier and mitochondrial DNA: evidence supporting the DNA pol gamma hypothesis", AIDS (London, England), vol. 20, no. 5, pp. 675-684.

Livak, K.J. & Schmittgen, T.D. 2001 “Analysis of relative gene expression data using real-time quantitative PCR and the $2^{(-\text{Delta Delta C(T)})}$ Method”. Methods. Vol. 25, No. 4:402–8.

Longley, M.J., Graziewicz, M.A., Bienstock, R.J. & Copeland, W.C. 2005, "Consequences of mutations in human DNA polymerase γ", Gene, vol. 354, no. 0, pp. 125-131.

Luigetti M, Sauchelli D, Primiano G, et al. 2016. “Peripheral neuropathy is a common manifestation of mitochondrial diseases: a single-centre experience.” European Journal of Neurology Vol. 23, pp. 1020–1027.

Malyarchuk, B., Perkova, M. & Derenko, M. 2011, "Polymorphism of 5′-promotor region of mitochondrial γ-DNA-polymerase gene in human populations", Molecular Biology, vol. 45, no. 5, pp. 852-853.

Maritz, J., Benatar, M., Dave, J.A., Harrison, T.B., Badri, M., Levitt, N.S. & Heckmann, J.M. 2010, "HIV neuropathy in South Africans: frequency, characteristics, and risk factors", Muscle & nerve, vol. 41, no. 5, pp. 599-606.

Menezes, C., Duarte, R., Dickens, C., DixPeek, T., Van Amsterdam, D., John, M., Ive, P., Maskew, M., Macphail, P. & 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, vol. 14, no. 4, pp. 217-225.

Migliore, L. & Coppedè, F. 2009, "Genetics, environmental factors and the emerging role of epigenetics in neurodegenerative diseases", Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, vol. 667, no. 1, pp. 82-97.

Miller, S., Dykes, D. & Polesky, H. 1988, "A simple salting out procedure for extracting DNA from human nucleated cells", Nucleic acids res, vol. 16, pp. 1215.

Montier, L.L.C., Deng, J.J. & Bai, Y. 2009, "Number matters: control of mammalian mitochondrial DNA copy number", Journal of Genetics and Genomics, vol. 36, no. 3, pp. 125-131.

Murray, C.J.L., Ortblad, K.F., Guinovart, C. *et al.* 2014 "Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013", The Lancet, vol. 384, no. 9947, pp. 1005-1070.

Nagiah, S., Phulukdaree, A. & Chuturgoon, A. 2015, "Mitochondrial and oxidative stress response in HepG2 cells following acute and prolonged exposure to antiretroviral drugs", Journal of cellular biochemistry, vol. 116, no. 9, pp. 1939-1946.

Naïmi, M., Bannwarth, S., Procaccio, V., Pouget, J., Desnuelle, C., Pellissier, J.F., Rötig, A., Munnich, A., Calvas, P., Richelme, C. and Jonveaux, P., 2006. Molecµlar analysis of ANT1, TWINKLE and POLG in patients with mµltiple deletions or depletion of mitochondrial DNA by a dHPLC-based assay. European journal of human genetics, Vol. *14*, no 8, pp.917-922.

Nattrass, N. 2006, "South Africa's "rollout" of highly active antiretroviral therapy: a critical assessment", Journal of acquired immune deficiency syndromes (1999), vol. 43, no. 5, pp. 618-623.

Nolan, D., Hammond, E., James, I., McKinnon, E. & Mallal, S. 2003, "Contribution of nucleoside-analogue reverse transcriptase inhibitor therapy to lipoatrophy from the population to the cellular level", Antiviral Therapy, vol. 8, no. 6, pp. 617-626.

Ojwach, D., Aldous, C., Kocheleff, P. & Sartorius, B. 2016, "Mutations of mtDNA polymerase-γ and hyperlactataemia in the HIV-infected Zulu population of South Africa", SAMJ: South African Medical Journal, vol. 106, no. 12, pp. 1254-1259.

Orrell, C., Harling, G., Lawn, S.D., Kaplan, R., McNally, M., Bekker, L. & Wood, R. 2007, "Conservation of first-line antiretroviral treatment regimen where therapeutic options are limited", Antiviral Therapy, vol. 12, no. 1, pp. 83.

Ortega, V.E. & Meyers, D.A. 2014, "Pharmacogenetics: Implications of race and ethnicity on defining genetic profiles for personalized medicine", Journal of Allergy and Clinical Immunology, vol. 133, no. 1, pp. 16-26.

Pettengill, E.A., Parmentier-Line, C. & Coleman, G.D. 2012, "Evaluation of qPCR reference genes in two genotypes of Populus for use in photoperiod and low-temperature studies", BMC Research Notes, vol. 5, no. 1, pp. 366.

Phillips, T.J., Cherry, C.L., Cox, S., Marshall, S.J. & Rice, A.S. 2010, "Pharmacological treatment of painful HIV-associated sensory neuropathy: a systematic review and meta-analysis of randomised controlled trials", PLoS One, vol. 5, no. 12, pp. e14433.

Ponamarev, M.V., Longley, M.J., Nguyen, D., Kunkel, T.A. & Copeland, W.C. 2002, "Active site mutation in DNA polymerase γ associated with progressive external ophthalmoplegia causes error-prone DNA synthesis", Journal of Biological Chemistry, vol. 277, no. 18, pp. 15225-15228.

Ponchel, F., Toomes, C., Bransfield, K., Leong, F.T., Douglas, S.H., Field, S.L., Bell, S.M., Combaret, V., Puisieux, A., Mighell, A.J. & Robinson, P.A. 2003. "Real-time PCR based on SYBR-Green I fluorescence: an alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions." *BMC Biotechnology*, Vol. *3*, *no.* 1, p.18.

Popanda, O., Seibold, P., Nikolov, I., Oakes, C.C., Burwinkel, B., Hausmann, S., Flesch-Janys, D., Plass, C., Chang-Claude, J. & Schmezer, P. 2013, "Germline variants of base excision repair genes and breast cancer: A polymorphism in DNA polymerase gamma modifies gene expression and breast cancer risk", International Journal of Cancer, vol. 132, no. 1, pp. 55-62.

Psarra, A.M.G. and Sekeris, C.E., 2011. Glucocorticoids induce mitochondrial gene transcription in HepG2 cells: role of the mitochondrial glucocorticoid receptor. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1813(10), pp.1814-1821.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I. & Daly, M.J. 2007, "PLINK: a tool set for whole-genome association and population-based linkage analyses", The American Journal of Human Genetics, vol. 81, no. 3, pp. 559-575.

Reddy, P.V.B., Gandhi, N., Samikkannu, T., Saiyed, Z., Agudelo, M., Yndart, A., Khatavkar, P. and Nair, M.P., 2012. HIV-1 gp120 induces antioxidant response

element-mediated expression in primary astrocytes: role in HIV associated neurocognitive disorder. Neurochemistry international, vol. 61, no. 5, pp.807-814.

Rooney, J.P., Ryde, I.T., Sanders, L.H., Howlett, E.H., Colton, M.D., Germ, K.E., Mayer, G.D., Greenamyre, J.T. & Meyer, J.N. 2015, "PCR based determination of mitochondrial DNA copy number in multiple species", Mitochondrial Regulation: Methods and Protocols, pp. 23-38.

Rötig, A. & Poulton, J. 2009, "Genetic causes of mitochondrial DNA depletion in humans", Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, vol. 1792, no. 12, pp. 1103-1108.

Scarpulla, R.C. 2008, "Transcriptional paradigms in mammalian mitochondrial biogenesis and function", Physiological Reviews, vol. 88, no. 2, pp. 611-638.

Schmittgen, T.D & Livak, K.J. "Analysing real-time PCR data by the comparative C(T) method." Nat Protocol. 2008; Vol. 3, No .6:1101–8.

Seidel-Rogol, B.L. & Shadel, G.S. 2002, "Modulation of mitochondrial transcription in response to mtDNA depletion and repletion in HeLa cells", Nucleic acids research, vol. 30, no. 9, pp. 1929-1934.

Shah, A., Vaidya, N.K., Bhat, H.K. & Kumar, A. 2016, "HIV-1 gp120 induces type1 programmed cell death through ER stress employing IRE1alpha, JNK and AP-1 pathway", Scientific reports, vol. 6, pp. 18929.

Shen, M., Zhang, L., Bonner, M.R., Liu, C., Li, G., Vermeulen, R., Dosemeci, M., Yin, S. & Lan, Q. 2008, "Association between mitochondrial DNA copy number, blood cell counts, and occupational benzene exposure", Environmental and molecular mutagenesis, vol. 49, no. 6, pp. 453-457.

Singh, A., Mitra, A.K., Rath, S.K. & Indian Genome Variation Consortium 2012, "Analysis of single-nucleotide polymorphisms of PEO1 gene in 55 ethnic groups of India", Chronicles of Young Scientists, vol. 3, no. 4, pp. 304.

Sinxadi, P.Z., Dave, J.A., Samuels, D.C., Heckmann, J.M., Maartens, G., Levitt, N.S., Wester, C.W., Haas, D.W. & Hµlgan, T. 2013, "Mitochondrial genomics

and antiretroviral therapy-associated metabolic complications in HIV-infected Black South Africans: a pilot study", AIDS Research and Human Retroviruses, vol. 29, no. 7, pp. 1031-1039.

Spitsina, E., Svetlova, G., Chudakova, D., Nikitin, A., Kurayeva, T., Strokov, I. & Nosikov, V. 2009, "Association of the POLG1 T (− 365) C, ANT1 G (− 25) A, and PEO1 G (− 605) T polymorphisms with diabetic polyneuropathy in patients with Type 1 diabetes mellitus", Molecular Biology, vol. 43, no. 2, pp. 348-352.

Stankov, M.V., Lucke, T., Das, A.M., Schmidt, R.E. & Behrens, G.M. 2010, "Mitochondrial DNA depletion and respiratory chain activity in primary human subcutaneous adipocytes treated with nucleoside analogue reverse transcriptase inhibitors", Antimicrobial Agents and Chemotherapy, vol. 54, no. 1, pp. 280-287.

Statistics South Africa, 2013. *www.statssa.gov.za/publications/P0318/P03182014.pdf*

Streatfield, P.K., Khan, W.A., Bhuiya, A., Hanifi, S.M., Alam, N., Millogo, O et al. 2014, "HIV/AIDS related mortality in Africa and Asia: evidence from INDEPTH health and demographic surveillance system sites", Global health action, vol. 7, pp. 25370.

Stumpf, J.D. & Copeland, W.C. 2011. “Mitochondrial DNA replication and disease: insights from DNA polymerase γ mutations.” *Cellular and Molecular Life Sciences*, Vol. *68, no.* 2, pp.219-233.

Suomalainen, A. & Isohanni, P. 2010, "Mitochondrial DNA depletion syndromes–many genes, common mechanisms", Neuromuscular Disorders, vol. 20, no. 7, pp. 429-437.

Svidnicki, M.C.C.C., Silva-Costa, S.M., Ramos, P.Z., dos Santos, N.Z.P., Martins, F.T.A., Castilho, A.M. and Sartorato, E.L., 2015. Screening of genetic alterations related to non-syndromic hearing loss using MassARRAY iPLEX® technology. BMC Medical Genetics, vol.16, no.1, pp.85.

Swan, E., Salem, R., Sandholm, N., Tarnow, L., Rossing, P., Lajer, M., Groop, P., Maxwell, A. & McKnight, A. 2015, "Genetic risk factors affecting mitochondrial function are associated with kidney disease in people with Type 1 diabetes", Diabetic Medicine, vol. 32, no. 8, pp. 1104-1109.

Taherzadeh-Fard, E., Saft, C., Akkad, D.A., Wieczorek, S., Haghikia, A., Chan, A., Epplen, J.T. & Arning, L. 2011, "PGC-1alpha downstream transcription factors NRF-1 and TFAM are genetic modifiers of Huntington disease", Molecular neurodegeneration, vol. 6, no. 1, pp. 1.

Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M. 2010, "A practical approach to RT-qPCR–Publishing data that conform to the MIQE guidelines. Methods." Vol 50, no 4: S1-S5.

The South African Antiretroviral Treatment Guidelines, 2013. www.sahivsoc.org/Files/2013%20ART%20Treatment%20Guidelines%20Final%2025%20March%202013%20corrected.pdf

Toyama, T., Iwase, H., Yamashita, H., Hara, Y., Omoto, Y., Sugiura, H., Zhang, Z. and Fujii, Y., 2003. Reduced expression of the Syk gene is correlated with poor prognosis in human breast cancer. Cancer letters, 189(1), pp.97-102.

UNAIDS.*www.unaids.org/sites/default/files/media.../AIDS_by_the_numbers_2015_en.pdf*

Urata, M., Koga-Wada, Y., Kayamori, Y. & 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, vol. 45, no. Pt 5, pp. 513-514.

Van Goethem, G., Dermaut, B., Löfgren, A., Martin, J. & Van Broeckhoven, C. 2001, "Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions", Nature Genetics, vol. 28, no. 3, pp. 211-212.

Van Oosterhout, J.J., Gardner, K., Mallewa, J., Kaunda, S., Kampira, E., Payne, B., Heyderman, R.S. & Chinnery, P. 2013, "Severe toxicity and polymerase-gamma

gene abnormalities in Malawian adults on stavudine-based antiretroviral therapy", Pharmacogenetics and genomics, vol. 23, no. 11, pp. 624-626.

Verma, S. & Simpson, D.M. 2007, "Peripheral neuropathy in HIV infection", Handbook of clinical neurology, vol. 85, pp. 129-137.

Wadley AL, Cherry CL, Price P, Kamerman PR, 2011. “HIV neuropathy risk factors and symptom characterization in Stavudine-exposed South Africans.” Journal of Pain Symptom Management, Vol. 41: pp.700–706.

Wadley, A.L., Lombard, Z., Cherry, C.L., Price, P. & Kamerman, P.R. 2012, "Analysis of a previously identified "pain-protective" haplotype and individual polymorphisms in the GCH1 gene in Africans with HIV-associated sensory neuropathy: a genetic association study", Journal of Acquired Immune Deficiency Syndromes (1999), vol. 60, no. 1, pp. 20-23.

Wadley, A.L., Lombard, Z., Cherry, C.L., Price, P. & 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, vol. 18, no. 1, pp. 94-96.

Wadley, A., Kamerman, P., Chew, C., Lombard, Z., Cherry, C. & Price, P. 2013, "A polymorphism in IL4 may associate with sensory neuropathy in African HIV patients", Molecular immunology, vol. 55, no. 3, pp. 197-199.

Wadley, A.L., Hendry, L.M., Kamerman, P.R., Chew, C.S., Price, P., Cherry, C.L. & Lombard, Z. 2014, "Role of TNF block genetic variants in HIV-associated sensory neuropathy in black Southern Africans", European Journal of Human Genetics.

Wang, W., Shen, P., Thiyagarajan, S., Lin, S., Palm, C., Horvath, R., Klopstock, T., Cutler, D., Pique, L., Schrijver, I., Davis, R.W., Mindrinos, M., Speed, T.P. & Scharfe, C. 2011, "Identification of rare DNA variants in mitochondrial disorders

with improved array-based sequencing", Nucleic Acids Research, vol. 39, no. 1, pp. 44-58.

Wiebe, L.A., Phillips, T.J., Li, J., Allen, J.A. & Shetty, K. 2011, "Pain in HIV: an evolving epidemic", The Journal of Pain, vol. 12, no. 6, pp. 619-624.

Workalemahu, T., Enquobahrie, D.A., Tadesse, M.G., Hevner, K., Gelaye, B., Sanchez, S.E. & Williams, M.A. 2017. “Genetic variations related to maternal whole blood mitochondrial DNA copy number: a genome-wide and candidate gene study.” The Journal of Maternal-Fetal & Neonatal Medicine, pp.1-7.

Xu, X., Sun, J., Chang, X., Wang, J., Luo, M., Wintergerst, K.A., Miao, L. & Cai, L. 2016, "Genetic variants of nuclear factor erythroid derived 2 like 2 associated with the complications in Han descents with type 2 diabetes mellitus of Northeast China", Journal of Cellular and Molecular Medicine, vol. 20, no. 11, pp. 2078-2088.

Yamanaka, H., Gatanaga, H., Kosalaraksa, P., Matsuoka-Aizawa, S., Takahashi, T., Kimura, S. & Oka, S. 2007, "Novel mutation of human DNA polymerase gamma associated with mitochondrial toxicity induced by anti-HIV treatment", The Journal of Infectious Diseases, vol. 195, no. 10, pp. 1419-1425.

Zhang, H. & Singh, K.K. 2014, "Global Genetic Determinants of Mitochondrial DNA Copy Number", PloS One, vol. 9, no. 8, pp. e105242.

APPENDICES

APPENDIX A

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

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Kamerman

CLEARANCE CERTIFICATE	**PROTOCOL NUMBER M080220**
PROJECT	Predictors of risk for developing antiretro-viral toxic neuropathy in patients of African ancestry infected with the HIV virus
INVESTIGATORS	Dr P Kamerman
DEPARTMENT	School of Physiology
DATE CONSIDERED	08.02.29
DECISION OF THE COMMITTEE* +	Approved subject to: Getting permission from relevant authorities

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE 08.03.10 **CHAIRPERSON** .....
(Professor P E Cleaton Jones)

*Guidelines for written 'informed consent' attached where applicable

cc: Supervisor : n/a

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University.
I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to a completion of a yearly progress report.**

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

APPENDIX C

Gene	SNP	Genotype	Frequency in SA
PINK1	rs10799655	TT	0.07
		TC	0.45
		CC	0.48
PINK1	rs650616	AA	0.17
		AG	0.5
		GG	0.33
PINK1	rs622525	AA	0.03
		AT	0.3
		TT	0.67
CCDC19	rs2501325	CC	0.03
		CA	0.21
		AA	0.76
ZFN648	rs7554182	TT	0.05
		TC	0.4
		CC	0.54
NRF2	rs1962142	AA	0
		AG	0.13
		GG	0.87
NRF2	rs35652124	CC	0.03
		CT	0.3
		TT	0.67
NRF2	rs2886162	AA	0.18
		AG	0.49
		GG	0.3
NRF3	rs4722585	AA	0.08
		AG	0.35
		GG	0.57
TOP1MT	rs7387720	GG	0.07
		GA	0.31
		AA	0.63
TOP1MT	rs2293925	AA	0
		AG	0.02
		GG	0.97

Gene	SNP	Genotype	Frequency in SA
	rs724037	GG	0.07
		GT	0.28
TOP1MT		TT	0.64
TOP1MT	rs11544484	CC	0.19
		CT	0.47
		TT	0.33
	rs2450772	AA	0.14
		AG	0.5
TOP1MT		GG	0.37
	rs11006126	CC	0.1
		CT	0.42
TFAM		TT	0.48
TFAM	rs2306604	AA	0.04
		AG	0.32
		GG	0.63
TFAM	rs1049432	TT	0.05
		TG	0.34
		GG	0.6
TFAM	rs11006132	GG	0.06
		GA	0.34
		AA	0.6
	rs17113613	AA	0.02
		AG	0.25
C10orf2		GG	0.73
C10orf2	rs3740485	CC	0.15
		CT	0.47
		TT	0.37
C10orf2	rs3740486	TT	0.15
		TC	0.48
		CC	0.38
C10orf2	rs11190787	GG	0.15
		GT	0.45
		TT	0.38

Gene	SNP	Genotype	Frequency in SA
POLG	rs2856268	CC	0.02
		CT	0.24
		TT	0.74
LITAF	rs13333308	CC	0.004
		CT	0.19
		TT	0.81
POLG2	rs9905016	TT	0.04
		TC	0.25
		CC	0.7
POLG2	rs2075551	CC	0.05
		CG	0.3
		GG	0.65