Investigation of the involvement of mitochondrial DNA variants in cardiometabolic disease: the SABPA study

M Pretorius
20196946
M.Sc

Thesis submitted for the degree *Philosophiae Doctor* in Biochemistry at the Potchefstroom Campus of the North-West University

Promoter:	Prof FH van der Westhuizen
Co-promoter:	Prof JL Elson
Assistant Promoter:	Prof L Malan

September 2017

this thesis is dedicated to my son, Aiden

thank you for being the most understanding 9 year old in the world,

I love you

Acknowledgements

I would like to thank the following people, without whom this thesis would not have been possible:

Prof Francois H. van der Westhuizen, my promoter, and extraordinary Prof Joanna L. Elson, my co-promoter, for their guidance and support, with endless patience. I am deeply grateful for the roles you have played in my life.

Prof Leoné Malan, my help-promoter, for her guidance and support.

Dr Etresia van Dyk, who worked hours well beyond what was expected, in order to sequence all the SABPA data.

Me Hayley van Dyk, who kindly performed all the practical work involved in the cybrid study.

Dr Eugene Engelbrecht, for thorough and timely language editing.

The present and past members of the Mitochondrial Research Laboratory, in particular Jaundrie Fourie, Karien Esterhuizen, Hayley van Dyk and Maryke Schoonen, who have all made my life easier in some professional and/or personal way during the past four years. I hope to repay this kindness in full.

All my friends and family, near and far, for their love, encouragements and support. I miss you all.

My parents, Arie and Louise Venter, and my Potchefstroom parents Marriëtte and J.C. Scholtz, for their unconditional love and support.

My siblings, Annette and Jay, for being Annette and Jay.

Zander Lindeque, for always having been there, in so many different ways.

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.

Table of contents

Abstract

Mitochondria are intricately involved in cell homeostatic and adaptive stress signalling pathways, and play a central role in cell differentiation, proliferation and death. Consequently, mitochondrial dysfunction has been implicated in a vast number of rare and common disease phenotypes. Mitochondrial DNA (mtDNA) encoded for 13 protein sub-units essential to mitochondrial function, as well as two subunits for mt-rRNA and 22 mt-tRNA molecules, required to translate and transcribe these proteins. As such, mtDNA variation, which could directly cause alterations in mitochondrial function and downstream processes, has been investigated as a possible risk factor in disease susceptibility, onset and progression. In this thesis, the role of mtDNA variation in cardiometabolic disease (CMD) is investigated. When mtDNA variation in common complex diseases such as CMD and other late onset and degenerative disorders are investigated, several approaches have been used to date, most notably the haplogroup association method. However, studies using these traditional methods have been plagued by inconsistencies, difficulties in replicating findings in other cohorts/populations, and contradicting reports. It is therefore clear that alternative approaches are needed in the field. A novel approach, the MutPred adjusted load hypothesis, is introduced in this thesis. This novel approach makes use of the MutPred scoring system to assign pathogenicity scores to non-synonymous mtDNA variants, which are then used to calculate a mutational load, a single statistical metric. The cumulative effect of several mildly deleterious variants can thus be measured in disease, using parametric statistical analyses. This new approach together with other more classic approaches were applied in a bi-ethnic South African cohort (N = 363) in this thesis. In addition, transmitochondrial cytoplasmic hybrids (cybrid) cells were utilised to investigate the impact of MutPred mutational loads on mitochondrial function. Using several investigative approaches, no significant associations between mtDNA variants and hypertension, hyperglycaemia, or indicators of inflammation and oxidative stress could be found ($P > 0.05$). However, in a preliminary study done in cybrid cells, several classifications of mtDNA variation, including MutPred mutational loads ($P < 0.01$), mtDNA variants with low MutPred scores ($P < 0.00001$), and relative mtDNA copy number ($P < 0.00001$) were shown to be significantly correlated with mitochondrial respiration rates. Further studies, investigating the underlying mechanisms of these relationships are warranted. Thus, while a role for MutPred mutational loads in CMD could not be found in the current cohort, a role for mtDNA variants in mitochondrial function in cybrid cells was found. In addition, it was demonstrated that the MutPred adjusted load hypothesis approach delivers more statistical power to studies when compared to haplogroup association studies, making it suitable for use even in moderately sizes cohorts. This approach should find wide application in the field, being especially useful for cohorts from multiple locations or with a variety of mtDNA lineages, where the traditional haplogroup association method has failed.

Keywords: mitochondrial DNA, MutPred, African, hypertension, hyperglycaemia, oxidative stress, inflammation, cybrids, population variant, rare variant.

Opsomming

Mitochondria is in noue betrokkenheid by sel homeostase asook aanpassings stresseinweë en speel 'n sentrale rol in sel-differensiasie, -vermeerdering en -dood. Gevolglik word die betrokkenheid van mitochondriale disfunksie by 'n groot aantal skaars en algemene siekte fenotipes vermoed. Mitochondriale DNA (mtDNA) kodeer vir 13 proteïen subeenhede wat noodsaaklik is mitochondrial funksie, asook twee subeenhede van mt-rRNA en 22 mt-tRNA molekules, wat nodig is vir transkripsie en translasie van hierdie proteïene. As sodanig is mtDNA variasie, wat direk tot veranderinge in mitochondriale funksie en stroomaf prosesse kan lei, al ondersoek as 'n moontlike risiko faktor in siekte vatbaarheid, aanvang en verloop. In hierdie tesis word die rol van mtDNA variasie in kardiometaboliese siektes (KMS) ondersoek. Wanneer mtDNA variasie in algemene komplekse siektes soos KMS en ander laat aanvang en degeneratiewe versteurings ondersoek word, word verskeie benaderings gebruik, veral die haplogroep assosiasie metode. Studies wat hierdie tradisionele metodes gebruik, word egter geteister deur teenstrydighede, onherhaalbaarheid van bevindinge in ander groepe / bevolkings, en kontrasterende verslae. Dit is dus duidelik dat alternatiewe benaderings in die veld nodig is. 'n Nuwe benadering, die MutPred aangepasde lading hipotese, word bekendgestel in hierdie tesis. Hierdie nuwe benadering maak gebruik van die MutPred tellingstelsel om patogenisiteitstellings toe te ken aan nie-sinonieme mtDNA variante, wat dan gebruik word om ‘n mutasie lading, as enkele statistiese term, te bereken. Dus kan die samewerkende effek van meer as een effens nadelige variant gemeet word in siektes, met behulp van parametriese statistiese toetse. Hierdie nuwe benadering, tesame met ander meer klassieke metodes, is toegepas in ‘n bi-etniese Suid-Afrikaanse kohort (N = 363) in hierdie tesis. Daarbenewens is transmitochondriale sitoplasmiese hibriede (cybrid) selle gebruik om die impak van MutPred mutasie ladings in mitochondriale funksie te ondersoek. Met behulp van verskeie ondersoekende benaderings, kon geen statisties betekenisvolle assosiasies tussen mtDNA variante en hipertensie, hiperglisemie, of aanwysers van inflammasie en oksidatiewe stres gevind word nie (P > 0.05). In 'n voorlopige studie wat in cybrid selle gedoen is, het verskeie klassifikasies van mtDNA-variasie, insluitende MutPred mutasie ladings (P <0.01), mtDNA-variante met lae MutPred-tellings (P <0.00001) en relatiewe mtDNA kopie getal (P < 0.00001) statisties betekenisvol korrelasies met mitochondriale respirasietempo's getoon. Verdere studies, wat die onderliggende meganismes van hierdie verhoudings ondersoek, is nodig. Dus, terwyl daar nie 'n rol vir MutPred mutasie ladings in KMS in die huidige kohort gevind kon word nie, is 'n rol vir mtDNA-variante in mitochondriale funksie in cybrid selle aangedui. Verder is ook getoon dat die MutPred aangepaste lading hipotese benadering meer statistiese krag verleen aan studies, veral in vergelyking met haplogroup assosiasie studies, wat dit ook geskik maak vir gebruik in selfs matige grootte kohorte. Hierdie benadering behoort wye toepassing in die veld te vind, en sal veral nuttig wees vir kohorte vanaf verskillende areas of met 'n verskeidenheid mtDNA geslagslyne, waar die tradisionele haplogroup assosiasie metode misluk het.

Sleutelwoorde: mitochondriale DNA, MutPred, Afrika, hoë bloeddruk, hoë bloedsuiker, oksidatiewe stres, inflammasie, bevolking variant, skaars variant.

List of tables

List of figures

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 7

Chapter 1: Preface

1.1 STUDY MOTIVATION AND RATIONALE

Selected glossary

Terms that could have ambiguous meaning or are new, are used throughout this thesis in the following way:

1) "mutation" here refers to any mtDNA variant that has been clinically proven, per set criteria as discussed in Chapter 2, to be pathogenic and cause disease. Mutations are thus "disease causing" variants.

2) "disease-associated" variants are mtDNA variants that have been associated with disease in some way, but do not necessarily at this point in time meet all the criteria set out to prove pathogenicity.

3) "MutPred scores" refer to pathogenicity scores which are assigned by the MutPred system (*mutpred.mutdb.org/about.html*) to any non-synonymous structural gene mtDNA variant, while "MutPred-scoring variant" refers to such variants. MutPred scores above 0.5 are considered an "actionable hypothesis" for pathogenicity, while scores above 0.75 are considered a "confident hypothesis".

4) "Yarham scores" refer to pathogenicity scores for mt-tRNA variants, as assigned by the system described in Yarham, et al., (2011). Variants are considered: neutral with a Yarham score below 7; possibly pathogenic with a score between 7 and 10; probably pathogenic with a score between 11 and 13, but with no function evidence such as single fibre, steady-state or cybrid analyses; and definitely pathogenic with a score above 10 and the inclusion of single fibre, steady-state or cybrid analyses evidence.

Cardiometabolic disease describes a collection of common complex disease phenotypes which present with cardiovascular, metabolic (such as blood glucose levels), inflammatory and other abnormalities. These in turn, serve as risk factors in the development and progression of cardiovascular disease (CVD) (Castro et al., 2003). Worldwide, CVD is the number one cause of morbidity and mortality. In Sub-Sahara Africa (SSA), while communicable diseases are still responsible for the most deaths each year, the steady urbanisation of previously rural populations brings with it behavioural and lifestyle changes that favour CVD development (Omboni et al., 2016; Yusuf et al., 2001). Studies in developed countries such as the USA, have shown that CVD and risk factors thereof are more severe in populations of African descent (African-American) than in their Caucasian counterparts (Okin et al., 2011). While not as abundant, studies of populations living within SSA have found that Africans develop CVD at an earlier age and with more severe outcomes than European populations (Moran et al., 2013; Owolabi et al., 2015). Importantly, hypertension is more prevalent in especially southern and eastern African populations, and disease outcome was shown to more often be haemorrhagic stroke, a consequence of hypertension (Mensah et al., 2015; Owolabi et al., 2015). To a lesser extent, ischemic heart disease, which is the leading disease outcome in Caucasians, is also present in African populations (Mensah et

al., 2015) and was also demonstrated in a Black South African male cohort (Malan et al., 2017). It is thus clear that more investigations are needed to explain the varied mechanisms and risk factors involved in CVD development and progression, in different population and gender groups. Several world-wide initiatives have been launched to address this growing epidemic in resource-limited SSA. There is no doubt that lifestyle and environmental factors contribute greatly to CVD development (Malan & Malan, 2016) and discrepancies between population groups. However, the contributions of genetic factors that might alter risk for, or progression of disease cannot be ignored. Unfortunately, genetic data on Africans in general, but especially on *well-phenotyped* African disease cohorts, are lacking (Mensah et al., 2015). These issues were repeatedly highlighted and discussed during several conferences in the past few years, including the Fourth Human Heredity and Health in Africa (H3Africa) consortium meeting, held in Cape Town in May in 2014 (Mensah et al., 2015); the 16th biennial congress of the Southern African Society for Human Genetics (SASHG), held in Pretoria in August 2015; and the Pharmacogenetics and Personalised Medicine conference held in Cape Town in April 2016. Notably for genetics in disease, the H3Africa initiative aims to address the above-mentioned data gaps and investigate the genetic factors involved in common complex and other disease, including CVD in SSA. The importance of establishing and integrating African based expert groups, with the input of international experts, for mitochondria-linked diseases was also a major impetus for a workshop held in Potchefstroom in 2014. The workshop was attended by several SA and UK based researchers and clinicians involved in mitochondrial disease and/or human genetics, to review the current understanding of mtDNA variation in disease in African populations (Meldau et al., 2016; van der Westhuizen et al., 2015).

One well-phenotyped cohort in South Africa is from the Sympathetic activity and Ambulatory Blood Pressure in Africans (SABPA) prospective cohort study, which was initiated in 2008-2009 (Malan et al., 2015). The overall aim was to investigate the contribution of a hyperactive sympathetic nervous system to cardiometabolic disease presentation in urbanised Africans (N = 200), compared to Caucasians (N = 209), all being teachers with similar socio-economic status and from the same geographical area. A follow-up study was undertaken in 2011-2012. The SABPA study was well-controlled and collected clinical measurements such as: objectively measured lifestyle factors, indices of obesity, physical activity, cardiovascular function, blood glucose levels, catecholamine metabolism, sex steroid hormones, inflammation, lipid profiles, HIV status, targeted and untargeted metabolomics data sets, oxidative stress status, and ageing. The cohort in total showed prevalence rates for hypertension (58.6%) and pre-diabetes (44.8%) (Malan et al., 2015). Based on 24 hour ambulatory blood pressure monitoring (ABPM) measurements alone, 66% of Black South African participants and 39.2% of Caucasian South African participants were classified as hypertensive (>130/80 mmHg) (Hamer et al., 2015). This cohort thus presented a unique opportunity to investigate genetic factors involved in CVD in a bi-ethnic cohort.

In addition to the role of mitochondrial energy metabolism in CVD, *mtDNA variants* have been implicated in several common complex diseases, including hypertension and diabetes type 2 (Achilli et al., 2011; Cardena et al., 2014; 2016; Chinnery et al., 2010; Lui et al., 2012). However, current approaches used when investigating mtDNA variation in relation to disease, most notably haplogroup association studies, suffer from limitations imposed by the complexities of mtDNA evolution and the consequent diversity in and between population groups. From the myriad of contradicting and irreproducible reports that have resulted from a decade of mitochondrial haplogroup association studies, it has become clear that new approaches are needed (Salas & Elson, 2015). The mutational load hypothesis was first proposed by Elson et al. (2006) and further refined here, with the use of MutPred pathogenicity scores for non-synonymous variants. Several approaches used to investigate mtDNA variation are described and applied in this thesis. The MutPred adjusted load hypothesis, however, offers an additional and novel approach to evaluate whether the presence of one or more mildly deleterious mtDNA variants could in some way contribute to disease progression or morbidity. Because this approach moves away from the use of common population variants, and highlights the role of presumably rarer variants, it seeks to overcome some of the difficulties faced by haplogroup association studies, such as population stratification. It also allows for the use of parametric statistics, decreasing the statistical burden of multiple testing. This results in increased statistical power, even for cohorts of moderate size. The MutPred adjusted load hypothesis therefore, at the initiation of this study, appeared to be a suitable additional and novel approach to investigate mtDNA variation in a moderately sized cohort.

1.2 PROBLEM STATEMENT

Africans living in rapidly urbanizing countries such as South Africa are facing an epidemic of vascular disease and hypertension with very limited information regarding the genetic factors contributing to this public health issue. The identification of causal factors, including genetic risk factors, is critical to encourage better lifestyle choices in those most at risk and allow for personalised advice. However, genetic data on SSA populations in relation to disease is lacking, and even in well-characterised studies on mtDNA variation in disease done mostly in Caucasians, current approaches have delivered many results that often include inconclusive or contradicting reports. It is therefore imperative to generate such data in an African cohort, in order to investigate disease mechanisms in the relevant genetic context using alternative investigative approaches. Targeting those most at risk would not only have a positive impact on the individual but also help resolve the major health crisis that CVDs represent to Southern Africans. This approach could also serve as a paradigm for other populations in developing countries.

1.3 AIMS AND OBJECTIVES

The aim of this study is to determine whether mtDNA variants play a significant role in the presentation or severity of cardiometabolic disease in a bi-ethnic cohort, using selected current and novel investigative approaches.

To achieve this aim, the following objectives were set for the study:

1. Generate and contribute complete mtDNA sequences for an under-represented population group (Sub-Saharan Africans), as well as a Caucasian group from the same geographical area
2. Test whether disease-associated variants or known pathogenic mutations are more frequently found in hypertensive groups than control groups
3. Test whether mitochondrial tRNA (mt-tRNA) and rRNA (mt-rRNA) variants are more abundant in hypertensive groups compared to control groups
4. Using a new approach (a modified mutational load hypothesis), investigate the possible role of non-synonymous protein coding mtDNA variants *in the presentation and severity of hypertension or hyperglycaemia*
5. Using a new approach (a modified mutational load hypothesis), investigate the possible role of non-synonymous protein coding mtDNA variants in *altered levels of oxidative stress and inflammation indicators*
6. Investigate the influence of selected mtDNA variants on mitochondrial respiration in cytoplasmic hybrid cells

1.4 STRUCTURE OF THESIS

As summarized below, this thesis is presented in eight chapters that include one peer-reviewed publication, one submitted manuscript, and one report being prepared for publication.

Chapter 2: Literature review

This chapter consists of a submitted manuscript in which the rationale for investigating mtDNA variation in the context of cardiovascular disease is given, followed by a discussion of current and alternative approaches used in such studies. Recommendations are then presented for future studies, especially in African populations.

- Submitted manuscript to *Cardiovascular Journal of Africa* (Manuscript number CVJSA-D-16-00139): **The aetiology of cardiovascular disease – a role for mitochondrial DNA?**
  Marianne Venter, Francois H. van der Westhuizen, Joanna L. Elson

Chapter 3: SABPA cohort description and mtDNA sequencing methods

This chapter gives a description and other relevant information of SABPA. The methods used to produce mtDNA sequencing data and identify mtDNA variants are also described, which addressed the first objective of this study. Methods that are specific to objectives 2 to 6 are described in each corresponding results chapter.

Results

Objectives 2 to 6 of this study are addressed in the following four chapters (Chapters 4 -7):

Chapter 4: Evaluating the presence of disease-associated mtDNA variants as reported on MITOMAP, and mt-tRNA and mt-rRNA variant frequency in a hypertension cohort

In this chapter, the second and third objectives are addressed.

Chapter 5: Evaluating the role of MutPred adjusted loads in hypertension and hyperglycaemia

This chapter consists of a peer-reviewed paper in which the fourth objective of this study is addressed.

- Published paper: **Using MutPred derived mtDNA load scores to evaluate mtDNA variation in hypertension and diabetes in a two-population cohort: The SABPA study**
  Marianne Venter, Leone Malan, Etresia van Dyk, Joanna L. Elson, Francois H. van der Westhuizen
  Published in *Journal of Genetics and Genomics*. 44 (2017): 139-149
  http://dx.doi.org/10.1016/j.jgg.2016.12.003

Chapter 6: Evaluating the role of MutPred adjusted loads in indicators of oxidative stress and inflammation

This chapter consists of a report that is being prepared for publication in which the fifth objective of this study is addressed.

- Paper in preparation for publication: **Mitochondrial DNA variation in oxidative stress and inflammation: the SABPA study**
  Marianne Venter, Leone Malan, Etresia van Dyk, Joanna L. Elson, Francois H. van der Westhuizen

Chapter 7: Utilising transmitochondrial cytoplasmic hybrid cells to test the MutPred mutational load hypothesis

In this chapter, the sixth objective of this study is addressed.

Chapter 8: Summary and conclusion

This chapter includes a summary and critical evaluation of the data presented in this thesis, followed by conclusions that can be made and recommendations for future studies.

References

References for Chapters 2 and 5 (submitted manuscript and published article) follow directly after each chapter and are in the formatting styles required by the respective journals. All other references used in this thesis are included in the "References" section and are in the APA style.

Appendices

Appendix A: SABPA study ethics approval
Appendix B: SABPA study participation information and consent form
Appendix C: Cybrids study consent form
Appendix D: Supplementary materials for article published in Journal of Genetics and Genomics (Chapter 5)

1.5 AUTHOR CONTRIBUTIONS

Submitted manuscript in Chapter 2: M Pretorius (neé Venter) was responsible for the literature review and manuscript writing. J.L. Elson and F.H. van der Westhuizen were involved in manuscript writing and supervision.

Peer reviewed paper in Chapter 5: As the custodian of the SABPA cohort, L. Malan advised on matters relating to the cohort description, input on cardiometabolic risk and intellectual input for the manuscript. E. van Dyk was the technician responsible for next generation sequencing on the Ion PGM™, and advised on data processing and writing of the sequencing section of *Methods* in this paper. J.L. Elson was involved in study design, supervision of statistical analyses, manuscript writing and supervision. F.H. van der Westhuizen was involved in study design, manuscript writing and supervision. M. Pretorius (neé Venter) was responsible for sample preparation for sequencing (including library building steps), data generation, mining and analyses, statistical analyses, and manuscript writing.

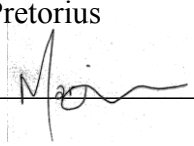
Paper in preparation for publication in Chapter 6: L. Malan was involved in study design and intellectual input. E. van Dyk was responsible for next generation sequencing on the Ion PGM™. J.L. Elson was involved in manuscript writing and supervision. F.H van der Westhuizen was involved in study design, manuscript writing and supervision. M. Pretorius (neé Venter) was involved in study design, and responsible for sample preparation for sequencing (including library building steps), data generation, mining and analyses, statistical analyses, and manuscript writing.

Cytoplasmic hybrid cell line study in Chapter 7: H.C. van Dyk was responsible for all laboratory practical work pertaining to cytoplasmic hybrid (cybrid) cell line production and maintenance, as well as Seahorse XF^e analyses and primary data generation. E.M. Schoeman and E. van Dyk were involved in the sample preparation and next generation on the Ion PGM™ sequencing for some cybrid cell lines included in this study. M. Pretorius was involved in the study design, as well as sample preparation and next generation sequencing of some cybrid cell lines included in this study, and also responsible for sequencing data generation, mining and analyses, Seahorse XF^e data analyses, statistical analyses and data interpretation.

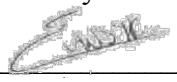
All persons involved signed the declarations on this page:

As co-author/researcher, I hereby approve and give consent that the mentioned articles and data can be used for the PhD of M. Pretorius. I declare that my role in the study, as indicated above, is a representation of my actual contribution.

Signature: M. Pretorius



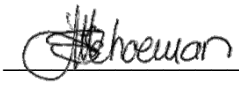
Signature: E. van Dyk



Signature:

H.C. van Dyk

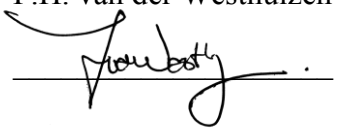
Signature: E.M. Schoeman



Signature: L. Malan

Signature: J.L. Elson

Signature: F.H. van der Westhuizen



Chapter 2: Literature review

Submitted for editorial review to: **Cardiovascular Journal of Africa**

2.1 TITLE

The aetiology of cardiovascular disease – a role for mitochondrial DNA?

2.2 AUTHORS AND AFFILIATIONS

Venter, Marianne[1]; Van der Westhuizen, Francois H. [1]; Elson, Joanna L. [1, 2]

1. Human Metabolomics, North-West University, Potchefstroom, South Africa
2. Institute of Genetic Medicine, Newcastle University, United Kingdom

Correspondence should be sent to: Marianne Venter, Human Metabolomics, North-West University, Potchefstroom, 2531, South Africa; Tel: +27 18 299 2318; Fax: +27 18 299 2477; Email: 20196946@nwu.ac.za

2.3 ABSTRACT

Cardiovascular disease (CVD) is a world-wide cause of mortality in humans, and on the rise in Africa. In this review, we discuss the putative role of mitochondrial dysfunction in the aetiology of CVD and consequently identify mitochondrial DNA (mtDNA) variation as a viable genetic risk factor to be considered. We then describe the contribution and pitfalls of several current approaches used when investigating mtDNA in relation to complex disease. We also propose an alternative approach, the adjusted mutational load hypothesis, which will have greater statistical power with cohorts of moderate size, and is less likely to be affected by population stratification. Therefore, it will address some of the shortcomings of the current haplogroup association approach. Finally, we discuss the unique challenges faced by studies done on African populations, and recommend the most viable methods to use when investigating mtDNA variation in CVD and other common complex disease.

Keywords: mitochondrial DNA; cardiovascular disease; MutPred; mutational load; African

2.3 INTRODUCTION

Cardiovascular disease (CVD) remains the main non-communicable cause of morbidity and mortality in humans [1]. While environmental factors and life style choices play a major role in CVDs, it is also recognized that genetic factors contribute significantly to the aetiology thereof. In this regard several studies, most recently genome wide association studies (GWAS), have contributed to identifying genetic loci involved in CVDs and their association with behavioural and biological risk factors [2][3][4][5][6][7]. Despite the numerous nuclear DNA (nDNA) variants identified, only a small portion of the heredity of CVDs can thus far be accounted for by variants discovered with GWAS studies [8]. For instance, the 46 loci identified for coronary artery disease (CAD) only account for about 6-13% of CAD hereditability [9][10][11]. The mitochondrion is the only other source of DNA apart from the nucleus. Mitochondrial DNA (mtDNA) encodes for 22 tRNAs, two rRNAs, and 13 polypeptides thought to be important in the catalytic cores of complexes I, III, IV and V of the oxidative phosphorylation (OXPHOS) system (Figure 1). In humans, mtDNA contains 16 569 bps and is double stranded [12]. Depending on the energy needs of a specific tissue, each cell can contain 100-1000s of copies of mtDNA [13]. MtDNA is maternally inherited and has a much higher mutation rate than nDNA, possibly 10-17 times higher [14]. Maternal inheritance results in a lack of bi-parental recombination, as such, the evolution of mtDNA is defined by the emergence of distinct lineages called haplogroups. Multi-copy makes possible a condition called heteroplasmy, where more than one genotype is present in the same cell/tissue/organism; homoplasmy then, is where all mtDNA copies carry the same allele. Notably, mtDNA is largely overlooked in GWASs, and could possibly contribute to the missing heredity of CVDs. Next, we will consider two main arguments on the possible role of mtDNA variants in CVDs.

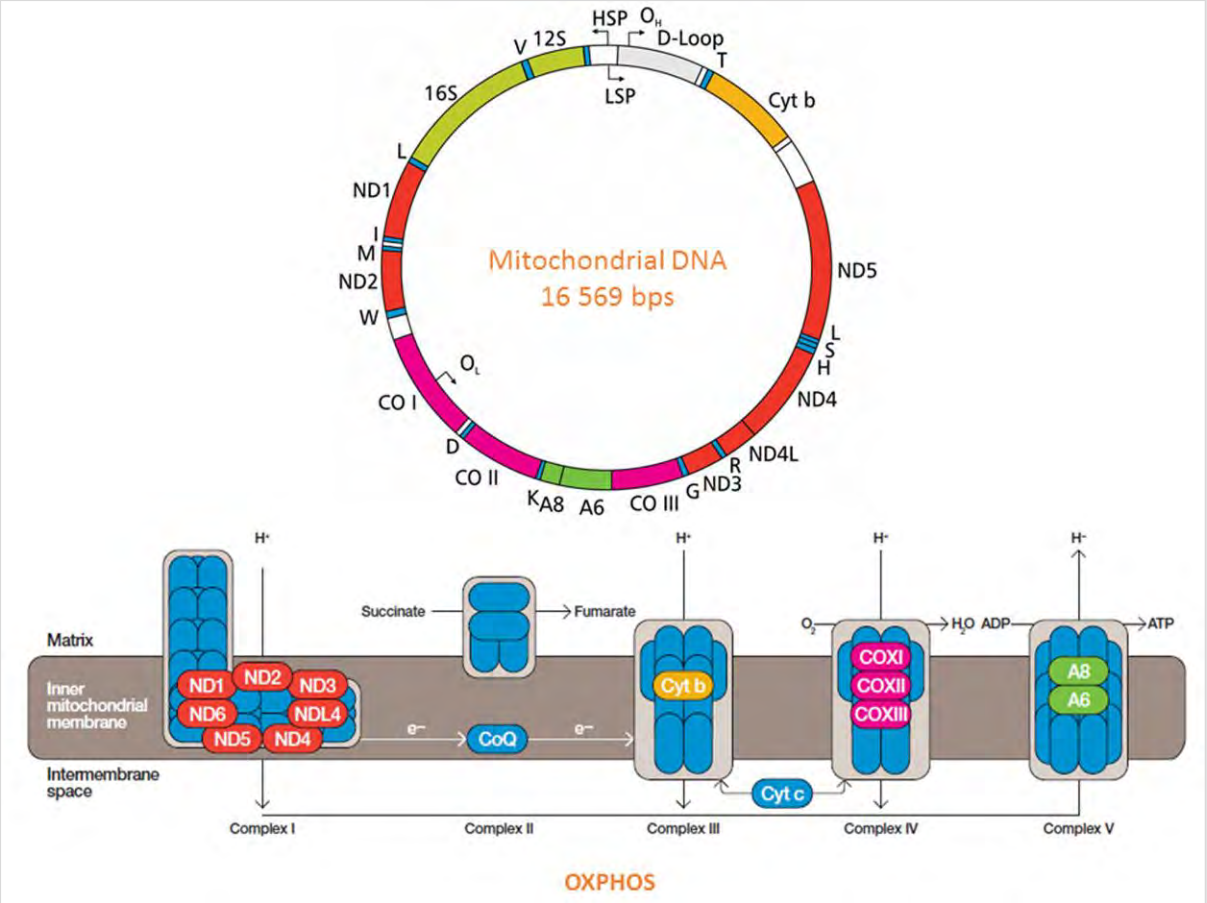


Figure 2.1: MtDNA molecule showing positions of mtDNA encoded genes, and the OXPHOS system showing positions of mtDNA encoded protein. mtDNA encodes for 22 tRNA and 2 rRNA molecules, as well as 13 polypeptide sub-units of the OXPHOS enzyme complexes, as indicated by colour. Enzyme complexes I-IV are involved in a series of redox reactions which transfer electrons from carriers nicotinamide adenine dinucleotide (NADH) and flavin adenine dinucleotide ($FADH_2$), to oxygen molecules. During these catalytically favourable reactions, H^+ -ions are pumped from the mitochondrial matrix into the mitochondrial intermembrane space, to create a proton-motor force across the inner-mitochondrial membrane. This force is used by Complex V to catalyse the phosphorylation of adenosine diphosphate (ADP) to adenosine triphosphate (ATP). Complex I: NADH dehydrogenase; complex II: succinate dehydrogenase; complex III: cytochrome c reductase; complex IV: cytochrome c oxidase; complex V: ATP synthase.

2.3.1 Mitochondrial dysfunction and mtDNA damage in vascular health

When considering mtDNA as a possible contributor in the aetiology of CVD, it should also be considered from a biological perspective: Much investigation has been conducted in an attempt to elucidate the risk factors and physiological mechanisms involved in the development of CVDs, such as sub clinical atherosclerosis, hypertension, cardiomyopathy and type 2 diabetes [15][16][17][18][19][20]. An important common feature in all these conditions is inflammation in some form or another (Figure 2). An inflammatory state is thought to be caused by oxidative stress, due to excessive levels of reactive oxygen species (ROS). ROS can be produced in several pathways, including by the enzymes such as

NADPH oxidase, nitric oxide synthase, and enzyme complexes of the electron transport chain (ETC) [21]. The general mechanism of ROS involvement in CVDs is ascribed to oxidative effects. For example, ROS contributes to atherosclerotic lesion formation by oxidising lipids, promoting vessel wall uptake of inflammatory cells, and enhancing proliferation and hypertrophy of vascular smooth muscle cells (VSMC) [21]. Several studies have shown increased levels of ROS in hypertensive humans and rats [16][22][23]. In cultured VSMCs for example, ROS has been shown to cause changes in cellular signalling pathways, favouring vasoconstriction [15]. A mechanism for this could be that ROS reduces nitric oxide (NO) bioavailability via quenching, impairing endothelial-mediated vasodilation [21][22][24]. However, ROS along with other factors of a dysfunctional mitochondrial energy metabolism (e.g. nucleotides, Ca^{2+}), also act as effectors of retrograde signalling and the so-called cell danger response [25][26][27].

Mitochondria are considered the major producers of ROS within the cell. In a recent article, Lopez-Armada et al. (2013) reviewed the role of mitochondrial dysfunction in the inflammatory response and consequently in the pathology of various diseases, including CVDs. The authors described how mitochondrial dysfunction might modulate inflammatory processes by activating redox-sensitive inflammatory pathways and the NLRP3 inflammasome. In the vasculature, these alterations lead to disturbed endothelial homeostasis, which has been implicated in the pathology of CVDs, such as atherosclerosis [18]. Indeed, some improvements in disease presentation of hypertension and diabetes have been observed in studies where chronic anti-oxidant treatment is applied [18][28][29]. Another mechanism by which inflammation might be altered by mitochondrial dysfunction is through the resultant release of mtDNA into the cytosol and circulation: because mtDNA is similar to bacterial DNA and not methylated [30], released mtDNA molecules are thought to induce an inflammatory state which contributes to atherosclerosis and other inflammatory diseases [31][32][33][34][35].

MtDNA damage has also been shown to promote atherosclerosis directly, in the absence of oxidative stress. In a study by Yu et al. (2013), VSMCs showed increased apoptosis and decreased proliferation in a proof-reading deficient $PolG^{-/-}/ApoE^{-/-}$ mouse model. Increased secretion of pro-inflammatory factors, tumour necrosis factor-α and interleukin-1β, were also reported and implicated in mtDNA release into the cytosol and subsequent activation of the inflammasome. The authors went on to test the applicableness of their findings in humans and concluded that an alternative mechanism for mtDNA defects mediate atherosclerosis development, independent of ROS: mtDNA defects lead to aberrant ECT function and consequently reduce ATP content in VSMCs, which then promotes apoptosis and inhibits cell proliferation, leading to increased atherosclerosis and risk of plaque rupture [36][37]. Plaque vulnerability is further promoted by mtDNA defects via monocyte cell death and the resultant increased release of inflammatory cytokine [38]. From these studies, it can be seen that mitochondrial dysfunction, possibly as a result of mtDNA variants or damage, can directly be implicated in mechanisms that encumber vascular health.

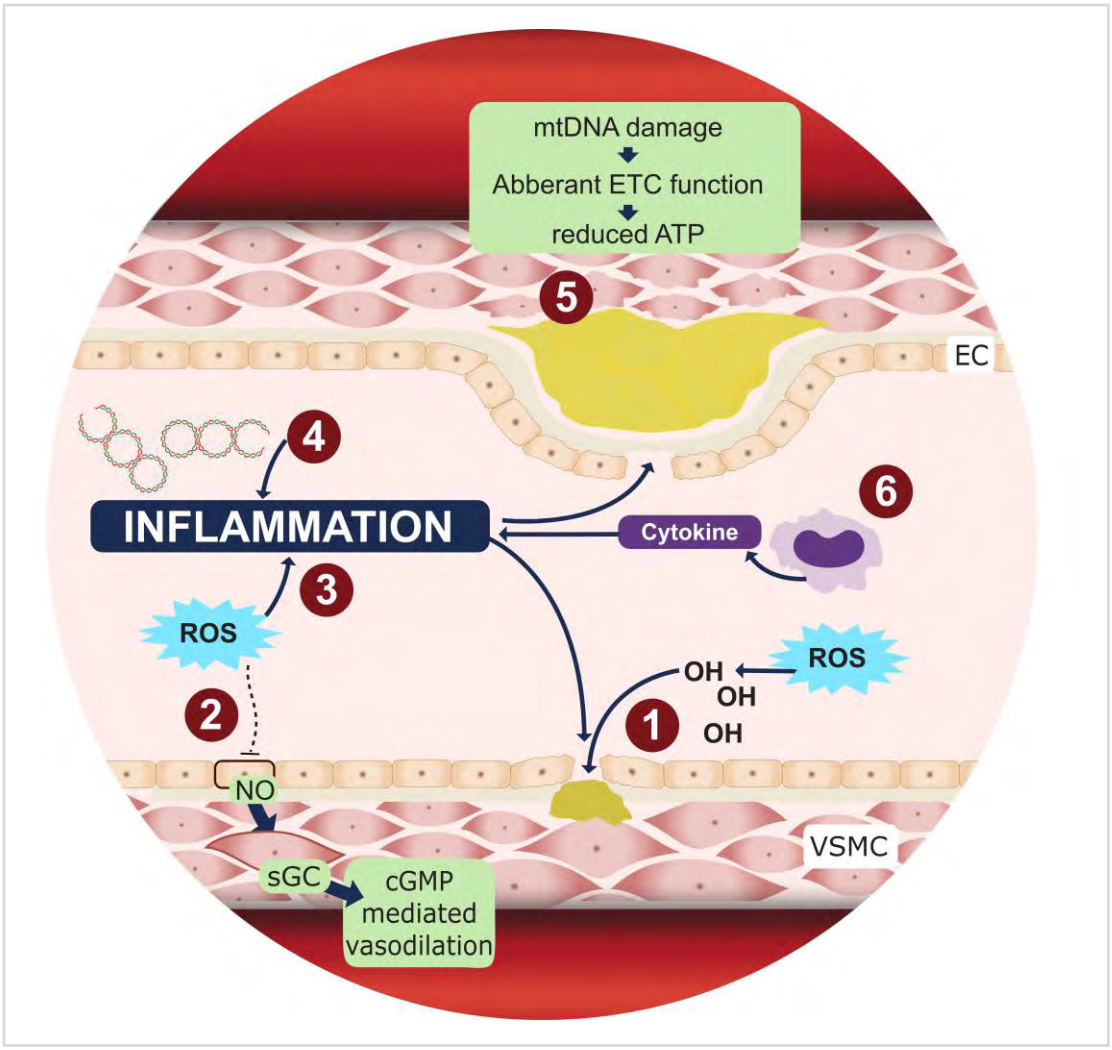


Figure 2.2: The role of mitochondrial dysfunction and mtDNA damage in vascular health. Mitochondrial dysfunction and mtDNA damage affect vascular health in several ways: 1. ROS aids in lesion formation by oxidising lipids, increasing the uptake of inflammatory cells into the vascular wall, and enhancing proliferation and hypertrophy in VSMC. 2. During endothelial dependant vasodilation, endothelial cells released NO activates soluble guanylyl cyclase in VSMC to produce cyclic GMP, signalling a vasodilation response. ROS inhibits this mechanism by quenching bioavailable NO molecules. 3. Endothelial homeostasis is disturbed and plaque formation promoted, when mitochondrial dysfunction leads to ROS formation and activates redox-sensitive inflammatory pathways. 4. Circulating cell free mtDNA is similar in structure to bacterial DNA and invokes an inflammatory response, contributing to atherosclerosis. 5. Independent from ROS formation, mtDNA damage leads to aberrant ETC function and reduced ATP levels in VSMC. When cell viability is compromised, apoptosis of VSMC occurs, accelerating plaque growth and decreasing plaque integrity. 6. Through the same mechanisms, apoptosis of monocytes occurs, releasing inflammatory cytokines, contributing to inflammation and consequently increasing plaque formation and vulnerability. ATP: adenosine triphosphate; cGMP: Cyclic guanosine monophosphate; EC: endothelial cell; ETC: electron transport chain; NO: nitric oxide; ROS: reactive oxygen species; sGC: soluble guanylyl cyclase; VSMC: Vascular smooth muscle cells.

2.3.2 MtDNA point mutations and cardiac involvement

Clinically proven mtDNA mutations are also an important cause of inherited disease [39]. To date, more than 250 deleterious point mutations and deletions of the mitochondrial genome have been clinically proven to be associated with certain disease phenotypes (www.mitomap.org). In several of these diseases, cardiovascular symptoms are an important part of the aetiology. Due to the very high levels of mtDNA population variation seen, both within and between human populations, the identification of mutations causing clinically manifesting disease prove to be a challenge, despite the small size of the mitochondrial chromosome. Initially, DiMauro and Schon (2001) had set specific criteria for defining the pathogenicity of mtDNA mutations. The list has subsequently been updated to include important methods such as functional testing and single fibre analysis, which can more specifically link genotype to phenotype [41][42]. Notably, a pathogenicity scoring system for mitochondrial tRNAs was devised by McFarland et al. (2004), and further refined by Yarham et al. (2011). Mitchell et al. (2006) also devised a pathogenicity scoring system using variants in complex I mtDNA genes, but this can be applied to any structural mtDNA mutation. A list of these criteria is given in Table 1.

Table 2.1: Criteria for defining the pathogenicity of mtDNA mutations

Criteria for pathogenicity of mtDNA mutations include:
• The mutation must only be present in patients and not controls • The mutation must be present in varied mitochondrial genetic backgrounds • The mutation must be the best mtDNA candidate variant to be pathogenic • The mutation must affect functionally important domains • Transfer of the mutated mtDNA to another cell line must be accompanied by transfer of the cellular or molecular defect • The mutation must not be a recognized, non-pathogenic SNP • The mutation must alter an area that is known to be highly conserved throughout evolution • The mutation must occur at varying levels within the cells (i.e. must be heteroplasmic) • A larger proportion of mutant mtDNA must correspond to a more severe phenotype • Single fiber PCR must be performed by comparing normal and abnormal fibers from muscle • The secondary structure of the tRNA molecule must also be taken into account when determining mt-tRNA mutation pathogenicity

Listed in Table 1, are criteria that need to be met, in order for a mtDNA mutation to be classified as "disease-causing", for either structural mtDNA or mt-tRNA mutations [40][41][42][43].

It should be noted that there are mtDNA mutations that are exceptions to all the "rules" in Table 1, and this was a critical motivation for algorithms or clinical scoring systems to help weigh the evidence that is presented for each mutation [43][44]. For a clinically proven mutation to manifest as a diseased phenotype, as in the case of primary mitochondrial disorders, the allele frequency (heteroplasmy) needs

to exceed a certain threshold, usually above 60%, referred to as the phenotypical threshold effect [45]. The biochemical threshold effect then, refers to the ability of the oxidative phosphorylation (OXPHOS) system to resist the metabolic expression of deficiencies therein [45][46]. These deficiencies may be caused by various factors involved in the expression and regulation of the OXHPOS complexes. There are many complexities to the expression of mtDNA mutations: a classic example is the mitochondrial tRNA mutation m.3243A>G, the most common of the mtDNA mutations causing mitochondrial disease. The m.3243A>G mutation can result in a vast array of clinical phenotypes affecting multiple systems within the body, causing two distinct clinical syndromes: maternally inherited diabetes and deafness (MIDD), and mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) syndrome in severe cases. Furthermore, the age of onset of m.3243A>G associated phenotypes span more than 50 years. The impact of several confounding factors, including heteroplasmy levels, remain unclear [47]. Another group of well-studied mutations are those that cause the disease Leber's hereditary optic neuropathy (LHON). In contrast to the m.3243A>G mutation, LHON has a tissue specific phenotype manifesting as bi-lateral blindness. Several mtDNA mutations have been implicated in LHON, while three of these mutations, namely m.3460G>A, m.11778G>A and m.14484T>C located in subunits ND1, ND4 and ND6 of complex I respectively, accounts for 90-95% of cases [48]. Unusually, these mutations can be detected as homoplasmic variants without exerting a phenotype. Rather, disease penetrance is significantly influenced by confounding factors such as gender and environment (clinical penetrance is increased to 93% in smoking men) [49] and mtDNA haplogroup background (haplogroup J, K and M7 increase risk of clinical penetrance) [50][51].

The heart has especially high energy needs and relies heavily on OXPHOS derived ATP, such that one third of cardiomyocyte volume consists of mitochondria [52]. Not surprisingly then, the myocardium is frequently affected in primary mitochondrial disorders [53]. In a retrospective review study by Yaplito-Lee et al. (2007), 33% of paediatric patients with definitive OXPHOS disorders had cardiac manifestations. Several mtDNA mutations (Figure 3 and Supplementary material) have also been shown to exhibit cardiac involvement, either as part of a multi-system syndrome (most frequently seen in MELAS), or as isolated occurrences, i.e. in the absence of associated CVDs or risk factors thereof [53][55][56]. Hypertrophic cardiomyopathy (hCM) and pulmonary artery hypertension (PAH) are the two phenotypes most commonly seen as isolated cardiac manifestations of primary mitochondrial disorders [53]. If clinically proven mtDNA mutations can directly lead to cardiac dysfunction, is it plausible to think that other mtDNA variants, such as population variants of mildly deleterious effect, might also lead to or alter severity/penetrance of complex cardiovascular disease phenotypes.

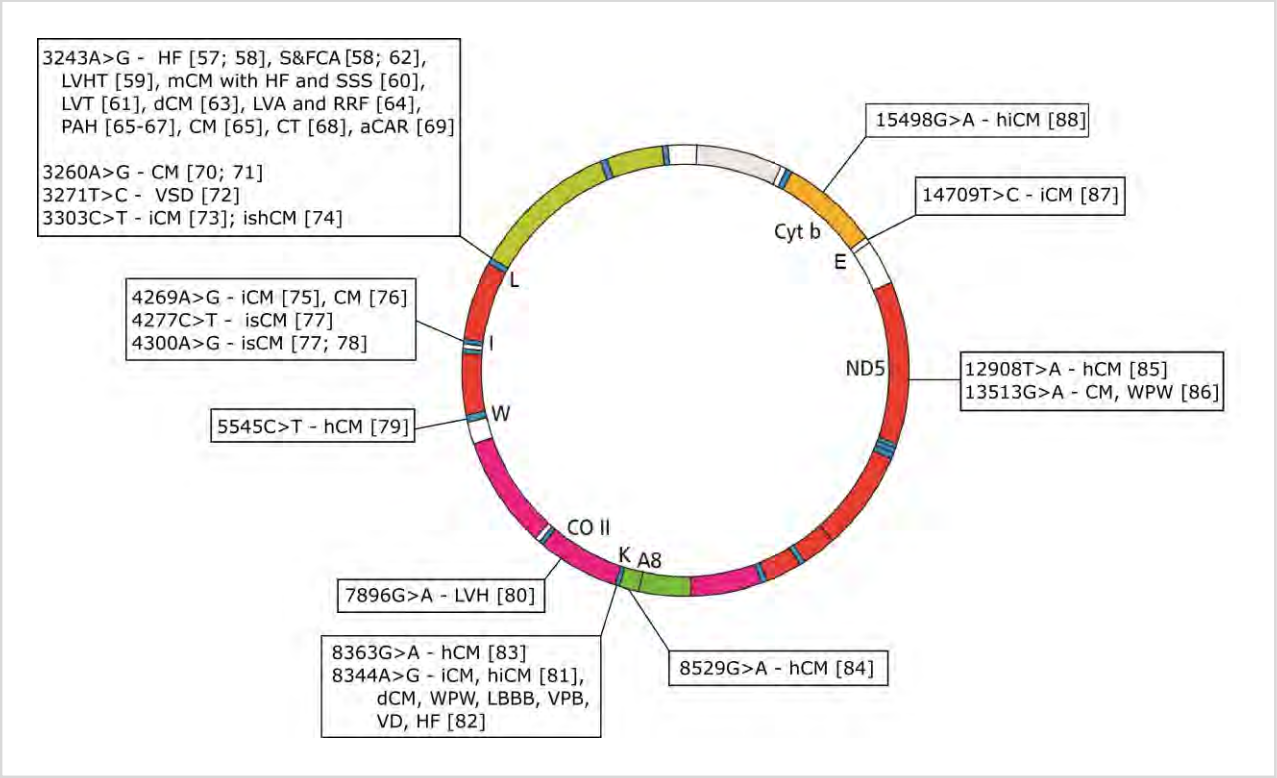


Figure 2.3: MtDNA morbidity map indicating clinically proven mtDNA mutations that present with syndromic or isolated cardiac involvement. aCAR: abnormal cardiac autonomic regulation; CM: cardiomyopathy; hCM: hypertrophic cardiomyopathy; dCM: dilated cardiomyopathy; HF: heart failure; hiCM: histiocytoid cardiomyopathy; iCM: infantile cardiomyopathy; ishCM: isolated hypertrophic cardiomyopathy; LBBB: left bundle branch block; LVA: left ventricle abnormalities; LVH: left ventricular hypertrophy; LVHT: left ventricular hypertrabeculation/noncompaction; mCM: mitochondrial cardiomyopathy; PAH: pulmonary artery hypertension; RRF: ragged red fibres; S&FCA: structural and functional cardiac abnormality; SSS: sick sinus syndrome; VD: ventricular dysfunction; VPB: ventricular premature beats; VSD: ventricle septal defect; WPW: Wolff–Parkinson–White syndrome. See Supplementary material for a detailed list of mutations, phenotype, references and pathogenicity scores as described in Mitchell et al. (2006) and Yarham et al. (2011).

From the substantial supportive evidence of mitochondrial involvement in cardiovascular disease, it is thus evident that genetics investigations on the aetiology of CVD should include consideration of mtDNA variation. In the following sections, we present a number of approaches (plus findings from such investigations) on how mtDNA variation is investigated/associate in/with disease, with a specific focus on the approaches more likely to show its putative contribution to the risk of CVD development.

2.4 CURRENT APPROACHES USED FOR INVESTIGATING MTDNA INVOLVEMENT IN DISEASE

2.4.1 Mitochondrial DNA copy number

MtDNA copy number can be used as an indicative marker of mitochondrial biogenesis, which is thought to increase in response to increased energy demands, such as exercise, but also as a compensatory method for mitochondrial dysfunction [89]. On the other hand, mtDNA copy number has been shown to decrease with aging [90] and has been significantly correlated with late onset diseases, such as Parkinson's disease [91][92]. As mentioned earlier, cell-free circulating mtDNA might also act as an inflammatory agent that contributes to CVDs [33]. Altered mtDNA copy number measured in peripheral blood cells have been shown to be associated with different complications of diabetes (diabetic retinopathy and diabetic nephropathy) [93][94]. Also, an association between telomere length and mtDNA copy number suggests a co-regulation mechanism for these two parameters, both of which are implicated in aging [95]. MtDNA depletion and impaired mitochondrial biogenesis have been shown to be a constant factor in the early stages of heart failure [96][97] and other disease thought to be related to aberrant ROS production [98]. While the exact mechanisms behind mtDNA content regulation are still unclear, it seems changes in either direction can be causative or indicative of disease [99]. Measurement of mtDNA copy number can be done accurately by real-time PCR methods, making this a useful approach for investigating the role of mitochondrial metabolism in disease phenotypes.

2.4.2 Common mtDNA population variants

MtDNA variants accumulated over time differ between population groups that have been separated for several thousand years. Consequently, distinct lineages (mtDNA haplogroups) can be drawn according to these sets of unique changes in mtDNA, referred to as common population variants. The full human mtDNA phylogeny can be accessed at *www.phylotree.org* [100]. Much of the variation seen in modern humans is to be found in the African haplogroups L0 to L6, but this variation has not been as fully described as the variation on other continents. European (eg. I, J, K, H, T, U, V, W, X) and Asian (eg. A, B, C, D, F, G) haplogroups fall within super haplogroups M and N, which in turn fall within L3. MtDNA haplogroup association studies therefore aim to associate these common mtDNA population variants with risk for various complex diseases, e.g. diabetes, hypertension or Parkinson's disease [101]. MtDNA background has been shown to correlate with the severity of cardiomyopathy caused by nDNA-encoded mitochondrial protein mutations [102], and increase the penetrance of LHON causing pathogenic mutations [50][51]. It has been proposed that mtDNA population variants could contribute to the adaptability of population groups to their environment, by altering mitochondrial enzyme function [103][104]. By analysing non-synonymous variants in 104 complete mtDNA sequences from across the globe, Mishmar et al. (2003) found that the *ATP6* and cytochrome *b* genes were particularly variable in

arctic and temperate zones respectively, leading them to believe that positive selection had taken place. Stressors, such as sudden changes in environment, could then influence the degree of disease susceptibility of these environmentally adapted population groups [105]. However, this hypothesis was contested by others who have shown that there are significant differences in the same measure in haplogroups from the same environment [106][107]. Additionally, Amo and Brand (2007) put forward evidence to suggest that certain bioenergetic parameters did not significantly differ between mitochondria from arctic vs tropical haplogroups. In contrast to the action of positive selection, the action of negative or purifying selection on mtDNA has been established for almost a decade [107][109]. One important point to consider, is that positive or directional selection could not have acted identically on all lineages, and as such would result in a different rate of accumulation of variants on haplogroup lineages, thus affecting our ability to time divergence events by the counting of mutational events between lineages. On the other hand, it is possible that negative or purifying selection could act evenly across lineages and not impact on our use of mtDNA as a molecular clock; the reliability of mtDNA as a molecular clock has been widely discussed [110].

Because of the central role that mitochondria play in cell signalling and apoptosis, mitochondria have been implicated in several age-related diseases, including Parkinson's disease, Alzheimer's disease, multiple sclerosis and psoriasis [101][111][112]. CVDs are also classified as late-onset disease, and mitochondria have also been implicated in CVDs. Consequently, haplogroup association studies on CVD phenotypes are plentiful – but, as will be revealed, also prone to pitfalls. Crispim et al. (2006) reported an association of European haplogroup cluster J/T, with insulin resistance and type 2 diabetes in a Caucasian-Brazilian cohort. On the other hand, Li et al. (2014) found no association between mtDNA variation and risk for developing diabetes, while Chinnery et al. (2007) found no association with type 2 diabetes and major European haplogroups in a large study using 897 cases and 1010 controls. Rather, Achilli et al. (2011) found that the risk for developing specific types of diabetes *complications* (disease outcome), is significantly associated with different mitochondrial haplogroups. Several mtDNA population variants in cytochrome c oxidase and NADH dehydrogenase subunit genes have been associated with body mass index (BMI) in adults [117]. In a very large study using a second cohort, Chinnery et al. (2010) found no significant associations between mtDNA haplogroups and ischaemic heart disease, hypertension, diabetes or metabolic syndrome, but did find a significant association of sub-haplogroup K with risk of cerebral ischaemic vascular effects. Thus, while some studies investigating phenotypes included in CVDs have reported results that support a role for mtDNA in CVD [116][117][119][120], there are also conflicting reports [115][118][121]. This is not only common in CVD related literature, but all areas where haplogroup association studies have been applied. This is an indication of the many difficulties that need to be overcome when considering mtDNA variation in the context of disease [122]. The unique way in which mtDNA is inherited (lack of bi-parental recombination), which results in the emergence of numerous unique haplogroups, contributes to the

complexity encountered when investigating mtDNA involvement in disease. Non-biological factors such as differences in statistical analysis approach [123]; difficulty in proper case and control matching; small effective population size, which results in a higher likelihood of population stratification; and insufficient cohort size [122], further undermine the consistency of these studies. Meta-analysis of data generated by several studies with overlapping phenotypes can be employed to overcome sample size difficulties, but these bring along challenges of their own, as independent studies have different goals/methods, and do not necessarily generate directly comparable datasets [101]. So, while haplogroup association studies might have fulfilled an important role in the ongoing pursuit of mtDNA variation involvement in disease, it is now well recognized that the field needs to consider alternative models.

2.4.3 Rare mtDNA population variants

It has been shown that negative or purifying selection plays a significant role in mtDNA evolution, with deleterious variants being removed from the population over time [107], and that the power of selection has been equally effective in all human lineages [124]. Consequently, rare mtDNA population variants are more likely to be mildly deleterious than common variants, as selection has had less time to remove them from the populations. Indeed, rare mtDNA variants have been linked to changes in CVDs and risk factors. In a study by Govindaraj et al. (2014), complete mtDNA analysis revealed ten non-synonymous variants present in hypertrophic cardiomyopathy patients, but not present in controls or on databases. Seven of these variants were classified as likely "pathogenic", using several online scoring tools such as PolyPhen-2, PMUT and PROVEAN, and were therefore thought to be involved in cardiomyopathy development. Rare variants m.5913G>A and m.3316G>A have both been suggested to be associated with increased fasting blood glucose levels, while m.5913G>A was shown to also be associated with increased blood pressure, in a selected Framingham heart study subset, all of whom were of European descent [7]. In addition, several rare mtDNA variants, such as m.3316G>A [7][125] have been implicated in diabetes mellitus, of which an up to date list can be found on *www.mitomap.org*. Another possibility is that the effect of an accumulation of mildly deleterious variants may only become clinically significant once a population is challenged by a rapid change of confounding factors, such as diet or other environmental factors (toxins) [126][127].

In conclusion, several approaches are currently in use for investigating the role of mtDNA in common complex disease. MtDNA copy number is an emerging approach that might become more prevalent in studies concerning CVDs as well. In terms of mtDNA variants, rare population variants have been linked to several disease phenotypes, including CVD related disease such as cardiomyopathy and diabetes mellitus, and might be found to be associated with other CVDs or risk factors such as hypertension. Rare population variants are more likely to be mildly deleterious [124], but might not have a high enough impact on their own to alter disease onset; rather, these variants might be more likely to

alter disease progression or outcome. For common population variants, several haplogroup association studies have been done in CVDs, but have also been marred by the challenges these type of studies face [122]. It seems then that an alternative approach to investigate the role of mtDNA variation in disease is needed when investigating common complex disease.

2.5 ALTERNATIVE APPROACH FOR INVESTIGATING MTDNA INVOLVEMENT IN DISEASE: THE ADJUSTED MUTATIONAL LOAD HYPOTHESIS

An alternative approach, the mutational load hypothesis, was put forward in Elson et al. (2006). Mutational load refers to the synergistic effect of several changes in e.g. a specific gene, or functionally related set of genes. It does not look for associations with a specific variant but rather a summative effect: while some mtDNA variants might be of negligible effect on their own, an increased mutational load might be associated with increased risk for a certain disease. MtDNA mutational loads can then be adjusted to reflect the position within the phylogeny, since there are large differences in the average number of common population variants between haplogroups. This approach can also further be modified to, for example, exclude low-impact variants, highlighting the role of likely deleterious functional variants. Determining the likely impact or pathogenicity of mtDNA variants can be achieved by using several computational pathogenicity predicting methods [128]. An example of such a method is the MutPred system, which assigns a MutPred score to any protein coding mtDNA variant, according to 14 gain/loss properties of protein structure and function [129]. The use of this system has been widely validated in the context of mtDNA studies [124], and performs better in an accuracy test when compared with several other methods [128]. Thus, the question can be asked whether individuals in the disease group are impacted by a combination of rare (mildly deleterious variants) or simply whether such variants are more common in the disease cohort than in the controls. The mutational load approach moves away from the study of haplogroups and looks at the collective effect of rare (or recent) variants, which are more likely to be deleterious. It distils the likely impact of a person's mtDNA variation into a single value on a continuous scale rather than a letter. Consequently, it will have more statistical power than conventional haplogroup association studies as more powerful parametric statistics can be applied, and fewer comparisons are required. As such, it offers an alternative method to explore the involvement of mtDNA variants in disease phenotypes, including diseases thought to be related to mitochondrial dysfunction, such as CVDs.

2.6 CONCLUSIONS: THE UNIQUE CHALLENGES FACED BY STUDIES IN AFRICAN POPULATIONS

While communicable diseases are still the leading causes of mortality in Sub-Saharan Africa (SSA), CVD is of particular a growing concern here, since the prevalence has risen most markedly in recent times, as more populations of developing countries becomes urbanised and are exposed to a diet and life style which increases risk factors for CVD [130]. Taking into account the many differences among ethnic groups in the onset and development of CVD [131][132], genomic investigations have also been used to investigate these disparities [130][133][134][135]. However, the number of well-powered genetic studies on CVDs in African populations or people of African descent is much lower than in European populations. As of yet, no conclusive nDNA genetic factor/s have been identified to help understand these disparities [136]. Current euro-centric reference panels used in GWAS studies to examine the involvement of population variants in disease, have been shown to be of limited use in even common SSA population groups [137]. This is indicative of the lack of African representation in our current databases. This lack extends to mtDNA as well: of the more than 30 000 mtDNA sequences available on GenBank, only 12% of these are of African lineages (L0-6). This bias in published data results in the resolution of the phylogenetic tree being much higher in the European branches (especially super-haplogroup N descendant) than in the African roots [138], despite greater diversity within the latter. Comparatively few studies have been done where the involvement of mtDNA variation in CVD has been considered [134][135][139][140][141]. Although of small size, one such study helps to highlight the challenges posed by these gaps in our current data: Ameh et al. (2011) could not find the tRNA mutation m.3243A>G in Nigerian diabetes type 2 patients, despite an association being previously reported in other European and Asian populations. This and other studies [142] illustrate the difficulty of extrapolating genetic risk factors for disease from one population group to the next, and the need for population specific studies.

In conclusion, SSA is facing a growing burden of CVD, while the discrepancies in onset and progression between different ethnicities are still poorly understood. Additionally, there are large data gaps when genetic studies on Africans are considered, especially for complex disease phenotypes. The unique genetic backgrounds of different populations also make it difficult to apply advances made in well-studied populations to understudied populations. While great efforts are being made to address these data gaps by initiatives such as the Human Heredity and Health in Africa (H3Africa) initiative [130], the Southern African Human Genome Programme, and the African Genome Variation Project [137], there is an urgent need for even more larger scale, African-specific investigations (which should also consider mtDNA variation) to be undertaken if we are to provide the necessary care to all vulnerable groups [143]. Realistically, for some time still, it is likely that studies in African populations will be hampered by financial and logistic/infrastructure difficulties [144], limiting the sizes thereof. Fortunately, these studies

can benefit from retrospective lessons we have learned thus far in other populations, highlighted in the above discussions. New studies could particularly benefit by asking better formulated questions, and using alternative approaches that aim to address the challenges associated with many of the classic approaches used, when the role of mtDNA in common disease is investigated.

2.7 KEY MESSAGES

- Cardiovascular disease (CVD) is a leading global cause for morbidity and mortality, and is on the rise in sub-Saharan Africa.
- Discrepancies in the onset and progression of CVDs exist between different ethnic and population groups, which nuclear genetic studies have so far failed to explain. Mitochondrial DNA (mtDNA) offers a viable alternative target for genetic studies concerning common complex disease.
- Many approaches can be taken to investigate the role of mtDNA in disease, but not all are suited for studies influenced by moderate cohort size or population stratification. The adjusted mutational load hypothesis offers an alternative approach, which could be of particular value for much-needed studies on CVDs in under-represented sub-Sahara African populations.

2.8 REFERENCES

1. Mensah GA. Descriptive epidemiology of cardiovascular risk factors and diabetes in Africa. *Prog. Cardiovasc. Dis.* 2013; 56: 240-250. http://dx.doi.org/10.1016/j.pcad.2013.10.014
2. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Hansen AT, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007; 316(5830): 1488-1491. doi:10.1126/science.1142447
3. Matarín M, Brown W, Scholz ,S Simón-Sánchez J, Fung H-C, Hernandez D, Gibbs J, Vrieze FD, Crews C, Britton A, Langefeld C, Brott T, Brown RD, Worrall BB, Frankel M, Silliman S, Case LD, Singleton A, Hardy JA, Rich SS, Meschia JF. A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. *Lancet Neurol* 2007; 6: 414-420. doi:10.1016/s1474-4422(07)70081-9
4. den Hoed M, Strawbridge RJ, Almgren P, Gustafsson S, Axelsson T, Engström G, de Faire U, Hedblad B, Humphries SE, Lindgren CM, Morris AP, Östling G, Syvänen AC, Tremoli E, Hamsten A, Ingelsson E, Melander O, Lind L. GWAS-identified loci for coronary heart disease are associated with intima-media thickness and plaque presence at the carotid artery bulb. *Atherosclerosis* 2015; 239: 304-310. http://dx.doi.org/10.1016/j.atherosclerosis.2015.01.032

5. Chen X, Kuja-Halkola R, Rahman I, Arpegärd J, Viktorin A, Karlsson R, Hagg S, Svensson P, Pedersen NL, Magnusson PK. Dominant genetic variation and missing heritability for human complex traits: Insights from twin versus genome-wide common SNP models. *Am. J. Hum. Genet.* 2015; 97: 708–714. http://dx.doi.org/10.1016/j.ajhg.2015.10.004
6. Arking DE, Chakravarti A. Understanding cardiovascular disease through the lens of genome-wide association studies. *Trends genet*. 2009; 25(9): 387-394. doi:10.1016/j.tig.2009.07.007
7. Liu C, Yang Q, Hwang S, Sun F, Johnson AD, Shirihai OS, Vasan RS, Levy D, Schwartz F. Association of genetic variation in the mitochondrial genome with blood pressure and metabolic traits. *Hypertension* 2012; 60: 949-956. doi: 10.1161/hypertensionaha.112.196519
8. Lotta L. Genome-wide association studies in atherothrombosis. *Eur. J. Intern. Med.* 2010; 21: 74-78. doi:10.1016/j.ejim.2009.11.003
9. Peden J, Farrall M. Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. *Hum. Mol. Genet.* 2011; 20(2): R198-R205. doi:10.1093/hmg/ddr384
10. The CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* 2013; 45(1): 25–33. doi: 10.1038/ng.2480
11. Smith JG, Newton-Cheh C. Genome-wide association studies of late-onset cardiovascular disease. *J Mol Cell Cardiol* 2015; 83: 131–141. http://dx.doi.org/10.1016/j.yjmcc.2015.04.004
12. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 1999; 23: 147. doi:10.1038/13779
13. Marín-García J, Akhmedov A. Mitochondrial dynamics and cell death in heart failure. *Heart Fail Rev* 2016; 21(2):123-136. doi: 10.1007/s10741-016-9530-2
14. Tuppen HA, Blakely EL, Turnbull DM, Taylor RW. Mitochondrial DNA mutations and human disease. *Biochim. Biophys. Acta.* 2010; 1797(2):113-128. doi: 10.1016/j.bbabio.2009.09.005
15. de Champlain J, Wu R, Girouard H, Karas M, EL Midaoui A, Laplante M, Wu L. Oxidative stress in hypertension. *Clin Exp Hypertens.* 2004; 26(7-8):593-601. PMID: 15702613
16. Salim S, Asghar M, Chugh G, Taneja M, Xia Z, Saha K. Oxidative stress: A potential recipe for anxiety hypertension and insulin resistance. *Brain res.* 2010; 359: 178-185. doi:10.1016/j.brainres.2010.08.093
17. Queisser N, Schupp N. Aldosterone oxidative stress and NF-kB activation in hypertension-related cardiovascular and renal diseases. *Free Radical Bio. Med.* 2012; 53: 314–327. http://dx.doi.org/10.1016/j.freeradbiomed.2012.05.011

18. López-Armada MJ, Riveiro-Naveira RR, Vaamonde-García C, Valcárcel-Ares MN. Mitochondrial dysfunction and the inflammatory response. *Mitochondrion* 2013; 13: 106–118. http://dx.doi.org/10.1016/j.mito.2013.01.003
19. Nakayama H, Otsu K. Translation of hemodynamic stress to sterile inflammation in the heart. *Trends Endocr. Met.* 2013; 24(11): 546-553. http://dx.doi.org/10.1016/j.tem.2013.06.004
20. van der Walt C, Malan L, Uys AS, Malan NT. Low grade inflammation and ECG left ventricular hypertrophy in urban African males: The SABPA study. *Heart lung circ.* 2013; 22(11): 924–929. doi: 10.1016/j.hlc.2013.03.075
21. Harrison DG, Gongora MC, Guzik TJ, Widder J. Oxidative stress and hypertension. *Journal of the American Society of Hypertension* 2007; 1(1): 30-44. doi:10.1016/j.jash.2006.11.006
22. Gönenç A, Hacışevk A, Tavil Y, Çengel A, Torun M. Oxidative stress in patients with essential hypertension: A comparison of dippers and non-dippers. *Eur. J. Intern. Med.* 2013; 24: 139–144. http://dx.doi.org/10.1016/j.ejim.2012.08.016
23. Yu EP, Bennett MR. The role of mitochondrial DNA damage in the development of atherosclerosis. *Free Radical Bio. Med.* 2016. doi: 10.1016/j.freeradbiomed.2016.06.011
24. Dantas AP, Franco M, d'Silva-Antonialli MM, Tostes RC, Fortes ZB, Nigro D, Carvalho MH. Gender differences in superoxide generation in microvessels of hypertensive rats: role of NAD(P)H-oxidase. *Cardiovasc. Res.* 2004; 61: 22– 29. doi:10.1016/j.cardiores.2003.10.010
25. Brière J, Chrétien D, Bénit P, Rustin P. Respiratory chain defects: what do we know for sure about their consequences in vivo? *Biochim. Biophys. Acta. (BBA)-Bioenergetics*. 2004; 1659(2): 172-177. http://dx.doi.org/10.1016/j.bbabio.2004.07.002
26. Reinecke F, Smeitink J, Van Der Westhuizen FH. OXPHOS gene expression and control in mitochondrial disorders. *Biochim. Biophys. Acta. BBA-Molecular Basis of Disease* 2009; 1792(12): 1113-1121. doi: 10.1016/j.bbadis.2009.04.003
27. Naviaux R. Metabolic features of the cell danger response. *Mitochondrion* 2014; 16: 7-17. doi: 10.1016/j.mito.2013.08.006
28. Stiefel P, Argüelles S, García S, Jiménez L, Aparicio R, Carneado J, Machado A, Ayala A. Effects of short-term supplementation with folic acid on different oxidative stress parameters in patients with hypertension. *Biochim. Biophys. Acta.* 2005; 1726: 152-159. doi:10.1016/j.bbagen.2005.07.014
29. Yamaguchi Y, Yamada K, Yoshikawa N, Nakamura K, Haginaka J, Kunitomo M. Corosolic acid prevents oxidative stress inflammation and hypertension in SHR/NDmcr-cp rats a model of metabolic syndrome. *Life Sci.* 2006; 79: 2474–2479. doi:10.1016/j.lfs.2006.08.007
30. Collins LV, Hajizadeh S, Holme E, Jonsson IM, Tarkowski A. Endogenously oxidized mitochondrial DNA induces in vivo and in vitro inflammatory responses. *J Leukoc Biol*. 2004; 75:995-1000

31. Zhou R, Yazdi A, Menu P, Tschopp J. A role for mitochondria in NLRP3 inflammasome activation. *Nature* 2011; 469(7329): 221-225. doi: 10.1038/nature09663
32. Nakahira K, Haspel J, Rathinam V, Lee S, Dolinay T, Lam H, Englert J, Rabinovitch M, Cernadas M, Kim H, Fitzgerald K, Ryter S, Choi A. Autophagy proteins regulate innate immune responses by inhibiting the release of mitochondrial DNA mediated by the NALP3 inflammasome. *Nat Immunol.* 2011; 12(3):222-230. doi: 10.1038/ni.1980
33. Oka T, Hikoso S, Yamaguchi O, Taneike M, Takeda T, Tamai T, Oyabu J, Murakawa T, Nakayama H, Nishida K, Akira S, Yamamoto A, Komuro I, Otsu K. Mitochondrial DNA that escapes from autophagy causes inflammation and heart failure. *Nature* 2012; 485(7397): 251-255. doi: 10.1038/nature10992
34. Shimada K, Crother T, Karlin J, Dagvadorj J, Chiba N, Chen S, Ramanujan V, Wolf A, Vergnes L, Ojcius D, Rentsendorj A, Vargas M, Guerrero C, Wang Y, Fitzgerald KA, Underhill DM, Town T, Arditi M. Oxidized mitochondrial DNA activates the NLRP3 inflammasome during apoptosis. *Immunity* 2012; 36(3): 401-414. doi: 10.1016/j.immuni.2012.01.009
35. West A, Khoury-Hanold W, Staron M, Tal M, Pineda C, Lang S, Bestwick M, Duguay B, Raimundo N, MacDuff D, Kaech S, Smiley J, Means RE, Iwasaki A, Shadel G. Mitochondrial DNA stress primes the antiviral innate immune response. *Nature* 2015; 520(7548): 553-537. doi:10.1038/nature14156
36. Yu E, Calvert P, Mercer J, Harrison J, Baker L, Figg N, Kumar S, Wang J, Hurst L, Obaid D, Logan A, West M, Clarke MC, Vidal-Puig A, Murphy MP, Bennett M. Mitochondrial DNA damage can promote atherosclerosis independently of reactive oxygen species through effects on smooth muscle cells and monocytes and correlates with higher-risk plaques in humans. *Circulation* 2013; 128: 702-712. doi: 10.1161/circulationaha.113.002271
37. Mercer JR. Mitochondrial bioenergetics and therapeutic intervention in cardiovascular disease. *Pharmacol. Therapeut.* 2014; 141(1):13-20. http://dx.doi.org/10.1016/j.pharmthera.2013.07.011
38. Yu EP, Bennett MR. Mitochondrial DNA damage and atherosclerosis. *Trends Endocr. Met.* 2014; 25(9): 481-487. http://dx.doi.org/10.1016/j.tem.2014.06.008
39. Lightowlers RN, Taylor RW, Turnbull DM. Mutations causing mitochondrial disease: What is new and what challenges remain? *Science* 2015; 349(6255): 1494-1499. doi: 10.1126/science.aac7516
40. DiMauro S, Schon E. Mitochondrial DNA mutations in human disease. *Am J Med Genet.* 2001; 106:18–26. doi: 10.1002/ajmg.1392
41. McFarland R, Elson JL, Taylor RW, Howell N, Turnbull DM. Assigning pathogenicity to mitochondrial tRNA mutations: when ‘‘definitely maybe’’ is not good enough. *Trends Genet.* 2004; 20:591–596. doi: 10.1016/j.tig.2004.09.014

42. Montoya J, López-Gallardo E, Díez-Sánchez C, López-Pérez MJ, Ruiz-Pesini E. 20 years of human mtDNA pathologic point mutations: Carefully reading the pathogenicity criteria. *Biochim. Biophys. Acta.* 2009; 1787:476–483. doi:10.1016/j.bbabio.2008.09.003
43. Yarham JW, Al-Dosary M, Blakely EL, Alston CL, Taylor RW, Elson JL, McFarland R. A comparative analysis approach to determining the pathogenicity of mitochondrial tRNA mutations. *Hum. Mutat.* 2011; 32(11):1319-1325. doi: 10.1002/humu.21575
44. Mitchell AL, Elson JL, Howell N, Taylor RW, Turnbull DM. Sequence variation in mitochondrial complex I genes: mutation or polymorphism? *J. Med. Genet.* 2006; 43:175–179. doi: 10.1136/jmg.2005.032474
45. Rossignol R, Faustin B, Rocher C, Malgat M, Mazat J, Letellier T. Mitochondrial threshold effects. *Biochem J.* 2003; 370(Pt 3):751-762. doi: 10.1042/bj20021594
46. Faustin B, Rossignol R, Rocher C, Bénard G, Malgat M, Letellier T. Mobilization of adenine nucleotide translocators as molecular bases of the biochemical threshold effect observed in mitochondrial diseases. *J. Biol. Chem.* 2004; 279(19): 20411–20421. doi 10.1074/jbc.m314259200
47. Picard M, Hirano M. Disentangling (epi)genetic and environmental contributions to the mitochondrial 3243A>G mutation phenotype: Phenotypic destiny in mitochondrial disease? *JAMA Neurol.* 2016; 73(8):923-925. doi: 10.1001/jamaneurol.2016.1676.
48. Kirches E. LHON: Mitochondrial mutations and more. *Curr. Genomics* 2011; 12: 44-54. doi: 10.2174/138920211794520150
49. Kirkman M, Yu-Wai-Man P, Korsten A, Leonhardt M, Dimitriadis K, De Coo I, Klopstock T, Chinnery PF. Gene-environment interactions in Leber hereditary optic neuropathy. *Brain* 2009; 132(Pt 9):2317-2326. doi: 10.1093/brain/awp158
50. Hudson G, Carelli V, Spruijt L, Gerards M, Mowbray C, Achilli A, Pyle A, Elson JL, Howell N, La Morgia C, Valentino M, Huoponen K, Chinnery PF. Clinical expression of Leber hereditary optic neuropathy is affected by the mitochondrial DNA-haplogroup background. *Am. J. Hum. Genet.* 2007; 81(2): 228-233. http://dx.doi.org/10.1086/519394
51. Ji Y, Zhang A, Jia X, Zhang Y, Xiao X, Li S, Guo X, Bandelt H, Zhang Q, Yao Y. Mitochondrial DNA haplogroups M7b1'2 and M8a affect clinical expression of Leber hereditary optic neuropathy in Chinese families with m.11778G>A mutation. *Am. J. Hum. Gen.* 2008; 83(6): 760-768. doi: 10.1016/j.ajhg.2008.11.002
52. Ong S-B, Kalkhoran S, Cabrera-Fuentes H, Hausenloy D. Mitochondrial fusion and fission proteins as novel therapeutic targets for treating cardiovascular disease. *Eur. J. Pharmacol.* 2015; 763:104-114. http://dx.doi.org/10.1016/j.ejphar.2015.04.056
53. Finsterer J, Kothari S. Cardiac manifestations of primary mitochondrial disorders. *Int. J. Cardiol.* 2014; 177: 754–763. http://dx.doi.org/10.1016/j.ijcard.2014.11.014

54. Yaplito-Lee J, Weintraub R, Jamsen K, Chow C, Thorburn D, Boneh A. Cardiac manifestations in oxidative phosphorylation disorders of childhood. *J. Pediatr.* 2007; 150(4): 407–411. doi: http://dx.doi.org/10.1016/j.jpeds.2006.12.047
55. Limongelli G, Masarone D, D'Alessandro R, Elliott PM. Mitochondrial diseases and the heart: an overview of molecular basis, diagnosis, treatment and clinical course. *Future cardiology*. 2012; 8(1):71-88. doi: 10.2217/fca.11.79.
56. Villar P, Bretón B, García-Pavía P, González-Páramos C, Blázquez A, Gómez-Bueno M, García-Silva T, García-Consuegra I, Martín MA, Garesse R, Bornstein B. Cardiac dysfunction in mitochondrial disease. *Circ. J.* 2013;77(11):2799-2806. doi: 10.1253/circj.CJ-13-0557
57. Yajima N, Yazaki Y, Yoshida K, Sano K, Takahashi W, Sasaki Y, Ikeda U. A case of mitochondrial cardiomyopathy with pericardial effusion evaluated by 99mTc-MIBI myocardial scintigraphy. *J. Nucl. Cardiol.* 2009; 16(6):989-994. doi:10.1007/s12350-009-9149-y
58. Malfatti E, Laforêt P, Jardel C, Stojkovic T, Behin A, Eymard B, Lombès A, Benmalek A, Bécane HM, Berber N, Meune C. High risk of severe cardiac adverse events in patients with mitochondrial m. 3243A> G mutation. *Neurology* 2013; 80(1):100-105. doi: http://dx.doi.org/10.1212/WNL.0b013e31827b1a2f
59. Finsterer J, Stöllberger C, Kopsa W. Noncompaction on cardiac MRI in a patient with nail-patella syndrome and mitochondriopathy. *Cardiology*. 2003; 100(1):48-49. doi:10.1159/000072393
60. Inamori M, Ishigami T, Takahashi N, Hibi K, Ashino K, Sumita S, Tamura K, Ochiai H, Umemura S, Ishii M, Tanaka S. A case of mitochondrial cardiomyopathy with heart failure, sick sinus syndrome and diabetes mellitus: mitochondrial DNA adenine-to-guanine transition at 3243 of mitochondrial tRNA (LEU)(UUR) gene. *J. Cardiol*. 1997; 30(6):341-347. PMID: 9436076
61. Majamaa-Voltti K, Peuhkurinen K, Kortelainen ML, Hassinen IE, Majamaa K. Cardiac abnormalities in patients with mitochondrial DNA mutation 3243A> G. *BMC cardiovascular disorders*. 2002; 2(1):1. doi: 10.1186/1471-2261-2-12
62. Hollingsworth KG, Gorman GS, Trenell MI, McFarland R, Taylor RW, Turnbull DM, MacGowan GA, Blamire AM, Chinnery PF. Cardiomyopathy is common in patients with the mitochondrial DNA m. 3243A> G mutation and correlates with mutation load. *Neuromuscular Disord.* 2012; 22(7):592-596. doi: http://dx.doi.org/10.1016/j.nmd.2012.03.001
63. Mima A, Shiota F, Matsubara T, Iehara N, Akagi T, Abe H, Nagai K, Matsuura M, Murakami T, Kishi S, Araoka T. An autopsy case of mitochondrial myopathy, encephalopathy, lactic

acidosis, and stroke-like episodes (MELAS) with intestinal bleeding in chronic renal failure. *Renal failure*. 2011; 33(6):622-625. http://dx.doi.org/10.3109/0886022X.2011.585730

64. Vydt TC, de Coo RF, Soliman OI, Folkert J, van Geuns RJ, Vletter WB, Schoonderwoerd K, van den Bosch BJ, Smeets HJ, Geleijnse ML. Cardiac involvement in adults with m. 3243A> G MELAS gene mutation. *Am. J. Cardiol.* 2007; 99(2):264-269. http://dx.doi.org/10.1016/j.amjcard.2006.07.089
65. Hung PC, Wang HS, Chung HT, Hwang MS, Ro LS. Pulmonary hypertension in a child with mitochondrial A3243G point mutation. *Brain and Development*. 2012; 34(10):866-868. http://dx.doi.org/10.1016/j.braindev.2012.02.011
66. Liu CH, Chang CH, Kuo HC, Ro LS, Liou CW, Wei YH, Huang CC. Prognosis of symptomatic patients with the A3243G mutation of mitochondrial DNA. *Journal of the Formosan Medical Association*. 2012; 111(9):489-494. http://dx.doi.org/10.1016/j.jfma.2011.06.014
67. Sproule DM, Dyme J, Coku J, de Vinck D, Rosenzweig E, Chung WK, De Vivo DC. Pulmonary artery hypertension in a child with MELAS due to a point mutation of the mitochondrial tRNA (Leu) gene (m. 3243A> G). *J. Inherit. Metab. Dis.* 2008; 31(3):497-503. doi:10.1007/s10545-007-0735-3
68. Wang W, Seak CJ, Liao SC, Chiu TF, Chen JC. Cardiac tamponade: a new complication in a patient with mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke like episodes. *Am. J. Emerg. Med.* 2008; 26(3):382-e1. doi: 10.1016/j.ajem.2007.05.027.
69. Majamaa-Voltti K, Majamaa K, Peuhkurinen K, Mäkikallio T, Huikuri H. Cardiovascular autonomic regulation in patients with 3243A> G mitochondrial DNA mutation. *Ann. Med.* 2004; 36(3):225-231. http://dx.doi.org/10.1080/07853890410028456
70. Sweeney MG, Brockington M, Weston MJ, Morgan-Hughes JA, Harding AE. Mitochondrial DNA transfer RNA mutation Leu (UUR) A→ G 3260: a second family with myopathy and cardiomyopathy. *QJM*. 1993; 86(7):435-438. doi: http://dx.doi.org/ 435-438. PMID: 8210299
71. Zeviani M, Gellera C, Antozzi C, Rimoldi M, Morandi L, Tiranti V, DiDonato S, Villani F. Maternally inherited myopathy and cardiomyopathy: association with mutation in mitochondrial DNA tRNALeu (UUR). *Lancet*. 1991; 338(8760):143-147. doi:10.1016/0140-6736(91)90136-D
72. Barišic N, Kleiner IM, Malcic I, Papa J, Boranic M. Spinal dysraphism associated with congenital heart disorder in a girl with MELAS syndrome and point mutation at mitochondrial DNA nucleotide 3271. *Croat Med J*. 2002;43(1):37-41. PMID: 11828557
73. Silvestri G, Santorelli FM, Shanske S, Whitley CB, Schimmenti LA, Smith SA, DiMauro S. A new mtDNA mutation in the tRNALeu (UUR) gene associated with maternally inherited cardiomyopathy. *Hum. mutat*. 1994; 3(1):37-43. doi: 10.1002/humu.1380030107

74. Palecek T, Tesarova M, Kuchynka P, Dytrych V, Elleder M, Hulkova H, Hansikova H, Honzik T, Zeman J, Linhart A. Hypertrophic cardiomyopathy due to the mitochondrial DNA mutation m. 3303C> T diagnosed in an adult male. *Int. Heart J.* 2012; 53(6):383-387. http://doi.org/10.1536/ihj.53.383
75. Hayashi J, Ohta S, Kagawa Y, Takai D, Miyabayashi S, Tada K, Fukushima H, Inui K, Okada S, Goto YI. Functional and morphological abnormalities of mitochondria in human cells containing mitochondrial DNA with pathogenic point mutations in tRNA genes. *J. Biol. Chem.* 1994; 269(29):19060-19066. PMID: 7518448
76. Taniike M, Fukushima H, Yanagihara I, Tsukamoto H, Tanaka J, Fujimura H, Nagai T, Sano T, Yamaoka K, Inui K, Okada S. Mitochondrial tRNAlle mutation in fatal cardiomyopathy. *Biochem. Bioph. Res. Co.* 1992; 186(1):47-53. doi:10.1016/s0006-291x(05)80773-9
77. Giordano C, Perli E, Orlandi M, Pisano A, Tuppen HA, He L, Ierinò R, Petruzziello L, Terzi A, Autore C, Petrozza V, Gallo P, Taylor RW and d'Amati G. Cardiomyopathies due to homoplasmic mitochondrial tRNA mutations: morphologic and molecular features. *Hum. Path.* 2013; 44(7):1262-1270. http://dx.doi.org/10.1016/j.humpath.2012.10.011
78. Casali C, Santorelli FM, Damati G, Bernucci P, DeBiase L, DiMauro S. A novel mtDNA point mutation in maternally inherited cardiomyopathy. *Biochem. Bioph. Res. Co.* 1995; 213(2):588-593. doi:10.1006/bbrc.1995.2172
79. Sacconi S, Salviati L, Nishigaki Y, Walker WF, Hernandez-Rosa E, Trevisson E, Delplace S, Desnuelle C, Shanske S, Hirano M, Schon EA, Bonilla E, De Vivo DC, DiMauro S and Davidson MM. A functionally dominant mitochondrial DNA mutation. *Hum. Mol. Genet.* 2008; 17(12):1814-1820. doi: 10.1093/hmg/ddn073
80. Villar P, Bretón B, García-Pavía P, González-Páramos C, Blázquez A, Gómez-Bueno M, García-Silva T, García-Consuegra I, Martín MA, Garesse R, Bornstein B and Gallardo ME. Cardiac dysfunction in mitochondrial disease. *Circ. J.* 2013;77(11):2799-2806. http://doi.org/10.1253/circj.CJ-13-0557
81. Vallance HD, Jeven G, Wallace DC, Brown MD. A case of sporadic infantile histiocytoid cardiomyopathy caused by the A8344G (MERRF) mitochondrial DNA mutation. *Pediatr. Cardiol.* 2004; 25(5):538-540. doi:10.1007/s00246-003-0446-y
82. Wahbi K, Larue S, Jardel C, Meune C, Stojkovic T, Ziegler F, Lombès A, Eymard B, Duboc D, Laforêt P. Cardiac involvement is frequent in patients with the m. 8344A> G mutation of mitochondrial DNA. *Neurology*. 2010; 74(8):674-677. http://dx.doi.org/10.1212/WNL.0b013e3181d0ccf4
83. Santorelli FM, Mak SC, El-Schahawi M, Casali C, Shanske S, Baram TZ, Madrid RE, DiMauro S. Maternally inherited cardiomyopathy and hearing loss associated with a novel mutation in the mitochondrial tRNA (Lys) gene (G8363A). *Am. J. Hum. Genet.* 1996; 58(5):933. PMCID: PMC1914622

84. Jonckheere AI, Hogeveen M, Nijtmans LG, van den Brand MA, Janssen AJ, Diepstra JH, van den Brandt FC, van den Heuvel LP, Hol FA, Hofste TG, Kapusta L, Dillmann U, Shamdeen MG, Smeitink JAM and Rodenburg, RJT. A novel mitochondrial ATP8 gene mutation in a patient with apical hypertrophic cardiomyopathy and neuropathy. *J. Med. Genet.* 2008; 45(3):129-133. doi:10.1136/jmg.2007.052084
85. Chamkha I, Alila-Fersi O, Mkaouar-Rebai E, Aloulou H, Kifagi C, Hachicha M, Fakhfakh F. A novel m. 12908T> A mutation in the mitochondrial ND5 gene in patient with infantile-onset Pompe disease. *Biochem. Bioph. Res. Co.* 2012; 429(1):31-38. http://dx.doi.org/10.1016/j.bbrc.2012.10.105
86. Wang SB, Weng WC, Lee NC, Hwu WL, Fan PC, Lee WT. Mutation of mitochondrial DNA G13513A presenting with Leigh syndrome, Wolff-Parkinson-White syndrome and cardiomyopathy. *Pediatrics & Neonatology*. 2008; 49(4):145-149. doi:10.1016/S1875-9572(08)60030-3
87. Van Hove JL, Freehauf C, Miyamoto S, Vladutiu GD, Pancrudo J, Bonilla E, Lovell MA, Mierau GW, Thomas JA, Shanske S. Infantile cardiomyopathy caused by the T14709C mutation in the mitochondrial tRNA glutamic acid gene. *Eur. J. Pediatr.* 2008; 167(7):771-776. doi:10.1007/s00431-007-0587-8
88. Andreu AL, Checcarelli N, Iwata S, Shanske S, Dimauro S. A missense mutation in the mitochondrial cytochrome b gene in a revisited case with histiocytoid cardiomyopathy. *Pediatr. Res.* 2000; 48(3):311-314. doi:10.1203/00006450-200009000-00008
89. Sahin E, DePinho R. Axis of ageing: telomeres p53 and mitochondria. *Nat. Rev. Mol. Cell Biol.* 2012; 13(6): 397–404. doi: 10.1038/nrm3352
90. Laderman K, Penny J, Mazzucchelli F, Bresolin N, Scarlato G, Attardi G. Aging-dependent functional alterations of mitochondrial DNA mtDNA from human fibroblasts transferred into mtDNA-less cells. *J. Biol. Chem.* 1996; 271: 15891-15897. PMID: 8663253
91. Pyle A, Anugrha H, Kurzawa-Akanbi M, Yarnall A, Burn D, Hudson G. Reduced mitochondrial DNA copy number is a biomarker of Parkinson's disease. *Neurobiol. Aging* 2016; 38:216.e7-216.e10. http://dx.doi.org/10.1016/j.neurobiolaging.2015.10.033
92. Gui Y-X, Xu Z-P, Lv W, Zhao J-J, Hu X-Y. Evidence for polymerase gamma POLG1 variation in reduced mitochondrial DNA copy number in Parkinson's disease. *Parkinsonism Relat. D.* 2015; 21:282-286. http://dx.doi.org/10.1016/j.parkreldis.2014.12.030
93. Malik AN, Parsade CK, Ajaz S, Crosby-Nwaobi R, Gnudi L, Czajka A, Sivaprasad S. Altered circulating mitochondrial DNA and increased inflammation in patients with diabetic retinopathy. *Diabetes Res. Clin. Pr.* 2015; 110(3):257-265. doi: 10.1016/j.diabres.2015.10.006
94. Czajka A, Ajaz S, Gnudi L, Parsade CK, Jones P, Reid F, Malik AN. Altered mitochondrial function, mitochondrial DNA and reduced metabolic flexibility in patients with diabetic

nephropathy. *EBioMedicine.* 2015; 2(6):499-512. doi: http://dx.doi.org/10.1016/j.ebiom.2015.04.002

95. Tyrka AR, Carpenter LL, Kao H-T, Porton B, Philipa NS, Ridout SJ, Ridout KK, Price LH. Association of telomere length and mitochondrial DNA copy number in a community sample of healthy adults. *Exp. Gerontol.* 2015; 66: 17–20. http://dx.doi.org/10.1016/j.exger.2015.04.002
96. Lagouge M, Larsson N. The role of mitochondrial DNA mutations and free radicals in disease and ageing. *J. Intern. Med.* 2013; 273(6): 529–543. doi: 10.1111/joim.12055
97. Pisano A, Cerbelli B, Perli E, Pelullo M, Bargelli V, Preziuso C, Mancini M, He L, Bates M, Lucena J, Della Monica P, Familiari G, Petrozza V, Nediani C, Taylor RW, d'Amati G, Giordano C. Impaired mitochondrial biogenesis is a common feature to myocardial hypertrophy and end-stage ischemic heart failure. *Cardiovasc. Pathol.* 2016; 25(2):103-112. http://dx.doi.org/10.1016/j.carpath.2015.09.009
98. Hao X-D, Chen P, Wang Y, Li S-X, Xie L-X. Mitochondrial DNA copy number but not haplogroup is associated with keratoconus in Han Chinese population. *Exp. Eye Res.* 2015; 132: 59-63. http://dx.doi.org/10.1016/j.exer.2015.01.016
99. Bersani FS, Morley C, Lindqvist D, Epel ES, Picard M, Yehuda R, Flory J, Bierer LM, Makotkine I, Abu-Amara D, Coy M, Reus VI, Lin J, Blackburn EH, Marmar C, Wolkowitz OM, Mellon SH. Mitochondrial DNA copy number is reduced in male combat veterans with PTSD. *Prog. Neuro-Psychoph.* 2016; 64: 10–17. http://dx.doi.org/10.1016/j.pnpbp.2015.06.012
100. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* 2009; 30(2):E386-394. doi: 10.1002/humu.20921
101. Hudson G, Gomez-Duran A, Wilson IJ, Chinnery PF. Recent mitochondrial DNA mutations increase the risk of developing common late-onset human diseases. *PLoS Genet* 2014; 10(5): e1004369. doi:10.1371/journal.pgen.1004369
102. Strauss KA, DuBiner L, Simon M, Zaragoza M, Sengupta PP, Li P, Narula N, Dreike S, Platt J, Procaccio V, Ortiz-Gonzalez XR, Puffenberger EG, Kelley RI, Morton DH, Narula J, Wallace DC. Severity of cardiomyopathy associated with adenine nucleotide translocator-1 deficiency correlates with mtDNA haplogroup. *Proc Natl Acad Sci USA* 2013; 110:3453-3458. doi: 10.1073/pnas.1300690110
103. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC. Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA.* 2003; 100(1):171-176. doi:10.1073/pnas.0136972100.
104. Ji F, Sharpley M, Derbeneva O, Alves L, Qian P, Wang Y, Chalkia D, Lvova M, Xu J, Yao W, Simon M, Platt J, Wallace DC. Mitochondrial DNA variant associated with Leber

hereditary optic neuropathy and high-altitude Tibetans. *Proc. Natl. Acad. Sci. USA*. 2012; 109(19):7391-7396. doi: 10.1086/519394

105. Wallace DC. Mitochondrial DNA variation in human radiation and disease. *Cell* 2015; 163(1): 33–38. doi:10.1016/j.cell.2015.08.067

106. Moilanen J, Finnilä S, Majamaa K. Lineage-specific selection in human mtDNA: lack of polymorphisms in a segment of MTND5 gene in haplogroup J. *Mol Biol. Evol*. 2003; 20: 2132–2142. doi: 10.1093/molbev/msg230

107. Elson JL, Turnbull DM, Howell N. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am. J. Hum. Genet.* 2004; 74: 229-238. doi: 10.1086/381505

108. Amo T, Brand M. Were inefficient mitochondrial haplogroups selected during migrations of modern humans? A test using modular kinetic analysis of coupling in mitochondria from cybrid cell lines. *Biochem. J.* 2007; 404: 345–351. doi:10.1042/bj20061609

109. Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Oefner PJ. The Role of selection in the evolution of human mitochondrial genomes. *Genetics* 2006; 172: 373–387. doi: 10.1534/genetics.105.043901

110. Howell N, Howell C, Elson JL. Molecular clock debate: Time dependency of molecular rate estimates for mtDNA: this is not the time for wishful thinking. *Heredity* 2008; 107-108. doi:10.1038/hdy.2008.52

111. Elson JL, Herrnstadt C, Preston G, Thal L, Morris CM, Edwardson JA, Beal MF, Turnbull DM, Howell N. Does the mitochondrial genome play a role in the etiology of Alzheimer's disease? *Hum. Genet.* 2006; 119: 241-254. doi 10.1007/s00439-005-0123-8

112. Maruszak A, Żekanowski C. Mitochondrial dysfunction and Alzheimer's disease. *Prog. Neuro-Psychoph.* 2011; 35(2): 320–330. doi: 10.1016/j.pnpbp.2010.07.004.

113. Crispim D, Canani LH, Gross JL, Tschiedel B, Souto KE, Roisenberg I. The European-specific mitochondrial cluster J/T could confer an increased risk of insulin-resistance and type 2 diabetes: an analysis of the m.4216T > C and m.4917A > G variants. *Ann Hum Genet*. 2006; 70(Pt 4):488-495. doi: 10.1111/j.1469-1809.2005.00249.x

114. Li S, Besenbacher S, Li Y, Kristiansen K, Grarup N, Albrechtsen A, Sparsø T, Korneliussen T, Hansen T, Wang J, Nielsen R, Pedersen O, Schierup MH. Variation and association to diabetes in 2000 full mtDNA sequences mined from an exome study in a Danish population. *Eur. J. Hum. Genet.* 2014; 22: 1040–1045. doi:10.1038/ejhg.2013.282

115. Chinnery PF, Mowbray C, Patel S, Elson JL, Sampson M, Hitman G, McCarthy M, Hattersley A, Walker M. Mitochondrial DNA haplogroups and type 2 diabetes: a study of 897

cases and 1010 controls. *J. Med. Genet.* 2007; 44(6):e80. http://dx.doi.org/10.1136/jmg.2007.048876

116. Achilli A, Olivieri A, Pala M, Hooshiar Kashani B, Carossa V, Perego UA, Gandini F, Santoro A, Battaglia V, Grugni V, Lancioni H, Sirolla C, Bonfigli AR, Cormio A, Boemi M, Testa I, Semino O, Ceriello A, Spazzafumo L, Gadaleta MN, Marra M, Testa R, Franceschi C, Torroni A. Mitochondrial DNA backgrounds might modulate diabetes complications rather than T2DM as a whole. *PloSone* 2011; 6(6): e21029. doi:10.1371/journal.pone.0021029

117. Flaquer A, Baumbach C, Kriebel J, Meitinger T, Peters A, Waldenberger M, Grallert H, Strauch K. Mitochondrial genetic variants identified to be associated with BMI in adults. *PLoSone* 2014; 9(8): e105116. doi:10.1371/journal.pone.0105116

118. Chinnery PF, Elliott HR, Syed A, Rothwell PM, Oxford Vascular Study. Mitochondrial DNA haplogroups and risk of transient ischaemic attack and ischaemic stroke: a genetic association study. *Lancet Neurol.* 2010; 9(5):498-503. doi: 10.1016/S1474-4422(10)70083-1

119. Elango S, Govindaraj P, Vishwanadha VP, Reddy G, Tamang R, Muthusami U, Kunnoth S, Koyilil VK, Lakshman M, Shanmugasundharam N, Singh L, Thangaraj K. Analysis of mitochondrial genome revealed a rare 50 bp deletion and substitutions in a family with hypertension. *Mitochondrion* 2011; 11: 878–885. doi:10.1016/j.mito.2011.07.002

120. Govindaraj P, Khanb NA, Rani B, Rani DS, Selvaraj P, Jyothi V, Bahl A, Narasimhane C, Rakshak D, Premkumar K, Khullar M, Thangaraj K. Mitochondrial DNA variations associated with hypertrophic cardiomyopathy. *Mitochondrion* 2014; 16:65–72. http://dx.doi.org/10.1016/j.mito.2013.10.006

121. Salas A, Elson JL. Raising doubts about the pathogenicity of mitochondrial DNA mutation m.3308T>C in left ventricular hypertraveculation/compactation. *Cardiology* 2011; 122(2):113-115. doi: 10.1159/000339348

122. Salas A, Elson JL. Mitochondrial DNA as a risk factor for false positives in case-control association studies. *J. Genet. Genomics* Apr 2015; 20;42(4):169-172. http://dx.doi.org/10.1016/j.jgg.2015.03.002

123. Samuels DC, Carothers AD, Horton R, Chinnery PF. The power to detect disease associated with mitochondrial DNA haplogroups. *Am. J. Hum. Genet.* 2006; 78(4):713-720. http://dx.doi.org/10.1086/502682

124. Pereira L, Soares P, Radivoiac P, Li B, Samuels D C. Comparing phylogeny and the predicted pathogenicity of protein variations reveals equal purifying selection across the global human mtDNA diversity. *Am. J. Hum. Genet.* 2011; 8;88(4):433-439. doi 10.1016/j.ajhg.2011.03.006.

125. Nakagawa Y, Ikegami H, Yamato E, Takekawa K, Fujisawa T, Hamada Y, Ueda H, Uchigata Y, Miki T, Kumahara Y Ogihara T. A new mitochondrial DNA mutation associated with non-insulin-dependent diabetes mellitus. *Biochim. Biophys. Res. Commun.* 1995; 209:664-668. doi:10.1006/bbrc.1995.1550

126. López-Gallardo E Iceta R Iglesias E Montoya J Ruiz-Pesini E. OXPHOS toxicogenomics and Parkinson's disease. *Mutat. Res.-Rev. Mutat.* 2011; 728(3):98-106. doi: 10.1016/j.mrrev.2011.06.004

127. López-Gallardo E Llobet L Emperador S Montoya J Ruiz-Pesini E. Effects of tributyltin chloride on cybrids with or without an ATP synthase pathologic mutation. *Environ. Health Persp.* 2016. http://dx.doi.org/10.1289/ehp182

128. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum. Mutat.* 2011; 32(4):358-368. doi: 10.1002/humu.21445

129. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 2009; 25(21): 2744-2750. doi:10.1093/bioinformatics/btp528

130. Owolabi MO, Mensah GA, Kimmel PA, Adu D, Ramsay M, Waddy SP, Ovbiagele B, Rabadan-Diehl C, Rasooly R, Akarolo-Anthony SA, Rotimi C. Understanding the rise in cardiovascular diseases in Africa: harmonising H3Africa genomic epidemiological teams and tools. *Cardiovas. J. Afr.* 2014; 25(3): 134-136. doi: 10.5830/cvja-2014-030

131. Sliwa K, Wilkinson D, Hansen C, Ntyintyane L, Tibazarwa K, Becker A, Stewart S. Spectrum of heart disease and risk factors in a black urban population in South Africa the Heart of Soweto Study: a cohort study. *Lancet* 2008; 371: 915–922. doi: 10.1016/S0140-6736(08)60417-1

132. Okin PM, Kjeldsen SE, Dahlöf B, Devereux RB. Racial differences in incident heart failure during antihypertensive therapy. *Circulation and Cardiovascular Qualitative Outcomes* 2011; 4:157-164. doi: 10.1161/circoutcomes.110.960112

133. Lai C-Q, Tucker KL, Choudhry S, Parnell LD, Mattei J, García-Bailo B, Beckman K, Burchard EG, Ordovás JM. Population admixture associated with disease prevalence in the Boston Puerto Rican health study. *Hum. Genet.* 2009; 125:199–209. doi 10.1007/s00439-008-0612-7

134. Cardena M, Ribeiro-dos-Santos A, Santos S, Mansur A, Pereira A, Fridman, C. Amerindian genetic ancestry is associated with higher survival rates compared to African and European ancestry in Brazilian patients with heart failure. *Int. J. Cardiol.* 2014; 176(2):527-528. http://dx.doi.org/10.1016/j.ijcard.2014.07.039

135. Cardena M, Ribeiro-Dos-Santos A, Santos S, Mansur A, Bernardez-Pereira S, Santos P, Pereira A, Fridman C. Mitochondrial and genomic ancestry are associated with etiology of

heart failure in Brazilian patients. *J. Hum. Hypertens*. 2016; 30(2):120-123. doi:10.1038/jhh.2015.39

136. Kaufman JS, Dolman L, Rushani D, Cooper RS. The contribution of genomic research to explaining racial disparities in cardiovascular disease: A systematic review. *Am. J. Epidemiol.* 2015; 181(7):464–472. doi: 10.1093/aje/kwu319

137. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, Karthikeyan S, Iles L, Pollard MO, Choudhury A, Ritchie GR, Xue Y, Kamali A. The African genome variation project shapes medical genetics in Africa. *Nature* 2015; 517: 327-332. doi:10.1038/nature13997

138. Cavadas B, Soares P, Camacho R, Brandao A, Costa MD, Fernandes V, Pereira JB, Rito T, Samuals D, Periera L. Fine time scaling of purifying selection on human nonsynonymous mtDNA mutations based on the worldwide population tree and mother-child pairs. *Hum. mutat*. 2014; 36(11): 1100-1111. doi: 10.1002/humu.22849

139. Khogali SS, Myosi BM, Beattie JM, McKenna WJ, Watkins H, Poulton JA. A common mitochondrial DNA variant associated with susceptibility to dilated cardiomyopathy in two different populations. *Lancet* 2001; 357: 1265-1267. doi:10.1016/S0140-67360004422-6

140. Robinson MT, Fischel-Ghodsian N, Fraser HS, Nicholson GD, Grim CM, Wilson DM, Wilson TW, Grim CE. Genetic influences on the increase in blood pressure with age in normotensice subjects in Barbados. *Ethnic. Dis.* 2004; 14: 57-63. PMID: 15002924

141. Ameh J, Godwin I, Obi I, Puepet F, Aminu B, Suleiman T. The search for mitochondrial tRNALeu(UUR) A3243G mutation among type 2 diabetes mellitus patients in the Nigerian population. *Afr. J. Biotechnol.* 2011; 10(62):13383-13389. doi: 10.5897/ajb11.1556

142. van der Walt EM, Smuts I, Taylor RW, Elson JL, Turnbull DM, Louw R, van der Westhuizen FH. Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease. *J. Hum. Genet.* 2012; 20(6):650-656. doi: 10.1038/ejhg.2011.262

143. van der Westhuizen FH, Sinxadi PZ, Dandara C, Smuts I, Riordan G, Meldau S, Malik AN, Sweeney MG, Tsai Y, Towers GW, Louw R, Gorman GS, Elson JL. Understanding the implications of mitochondrial DNA variation in the health of Black Southern African populations: The 2014 Workshop. *Hum. Mutat.* 2015; 36(5):569-571. doi: 10.1002/humu.22789

144. Forero DA, Wonkam A, Wang W, Laissue P, López-Correa C, Fernández-López JC, Mugasimangalam R, Perry G. Current needs for human and medical genomics research infrastructure in low and middle income countries. *J. Med. Genet*. 2016 ;53(7):438-40. doi: 10.1136/jmedgenet-2015-103631

Chapter 3: Methods

In this chapter, the Sympathetic Ambulatory Blood Pressure in Africans (SABPA) cohort is described, as well as the methods that were used to obtain complete mitochondrial DNA (mtDNA) next generation sequences from blood samples (first objective of this study). These general methods apply to all instances throughout the study where sequencing data were produced. Since the data processing and statistical analyses used differed for each of the remaining objectives, these specific methodological approaches are described separately where relevant in Chapters 4 to 7. This also applies to specific methods used for the cybrid studies conducted.

3.1 SABPA COHORT

The SABPA prospective study was set up to investigate several aspects of cardiometabolic disease in a Black South African population when compared to a geographically matched population of European descent. To ensure that all participants were of similar socio-economic status, 2170 teachers enrolled in schools in the Dr Kenneth Kaunda Education District (North-west Province, South Africa) were invited to participate in the study. During recruitment and before consent, study participants were informed about the objectives and procedures of the study, including investigations of genes involved in cardiometabolic disease. The study complied with criteria on human research set by the Helsinki declaration (2004) and was approved by the Ethics Committee of the North-West University (NWU 00036-07-S6). Exclusion criteria were: pregnancy, lactation, users of α- and β-blockers, psychotropic substance abuse, blood donors or those who had been vaccinated in the previous three months, and those that have a tympanum temperature above 37.5 °C. After screening, 409 eligible responders were enrolled, of which 200 were Black South Africans (101 males and 99 females), and 209 were Caucasians South Africans (101 males and 108 females). Baseline measurements were taken from 2008-2009, with a 3-year follow-up completed during 2011-2012. Clinical and biochemical measurements included those of hypothalamic-pituitary-adrenal (HPA) and sympathy-adrenal-medullary (SAM) axes; renin-angiotensin-aldosterone system; ambulatory blood pressure and electrocardiogram; neurotrophines; renal function; metabolism; and pro-inflammatory, pro-thrombotic, functional and structural endothelium (Malan et al., 2015).

For the current study, only baseline measurements were used. For some participants, blood samples were not available for next generation sequencing. Thus, for the 409 SABPA participants that are enrolled in the study, mtDNA sequences were obtained for only 363 participants (100 Black South African males, 94 Black South African females, 75 Caucasian South African males and 94 Caucasian South African females).

Some phenotypical data on the SABPA cohort important for this study is given in Table 3.1.

Table 3.1: Phenotypical data on SABPA cohort

Parameter	Black males	Black females	Caucasian males	Caucasian females
Number of participants (*n*)	100	94	75	94
Age (years)	43.2 ± 8.17	45.7 ± 7.86	45.4 ± 10.5	44.5 ± 10.7
Body mass index	27.6 ± 5.77	32.9 ± 7.23	29.0 ± 5.33	26.2 ± 5.74
HbA1c (%)	6.23 ± 1.23	5.90 ± 1.14	5.66 ± 0.49	5.36 ± 0.29
24 h systolic ABPM (mmHg)	137.7 ± 16.0	128.6 ± 15.2	128.1 ± 10.6	120.7 ± 12.7
25 h diastolic ABPM (mmHg)	87.9 ± 10.7	78.8 ± 8.50	79.7 ± 7.72	73.8 ± 7.89

For each parameter, the group mean value is given ± standard deviation. Body mass index is a measure of body fat based on height and weight. HbA1c, glycated haemoglobin is measured to calculate the three month average plasma glucose concentration. 24 h ABPM, 24 hour ambulatory blood pressure monitoring.

3.2 HYPERTENSION AND HYPERGLYCAEMIA IN THE SABPA COHORT

The current international guidelines for classification of hypertension provide thresholds of >130/80 mmHg for 24 hour ABPM (Mancia et al., 2013). However, Hermida et al. (2013) reported that ABPM measurements significantly differ between males and females. The authors propose that these current guidelines are only applicable in males, and that the derived outcome-based thresholds of day and night APBM measurements were 10/5 mmHg lower in females compared to males. Taking this into consideration, a second, but conservative threshold of 125/75 mmHg was chosen for hypertension classifications females in this study. Thus, a hypertensive classification was assigned to males with 24 h systolic ABPM >130 mmHg and/or 24 h diastolic ABPM >80 mmHg, and females with 24 h systolic ABPM >125 mmHg and/or 24 h diastolic ABPM >75 mmHg (Hermida et al., 2013). Glycated haemoglobin (HbA1c) measurements expressed as a percentage, which are measured to calculate the three-month average plasma glucose concentrations, were used to classify participants as having either optimal blood glucose levels (HbA1c below 5.7%), falling within pre-diabetic ranges (HbA1c from 5.7% to 6.4%) or diabetic ranges (HbA1c above 6.4%) (American Diabetes Association, 2010). Alternatively, participants were classified as having optimal blood glucose levels or being pre-diabetic, by applying the clinically-used threshold value of HbA1c = 5.7%. Incidences in the SABPA participants used in this study of hypertension and hyperglycaemia according to these criteria, are given in Table 3.2.

Table 3.2: Incidences of hypertension and hyperglycaemia in the SABPA cohort

SABPA cohort	*N*	% of participants that are:				
		Normotensive	Hypertensive	Non-diabetic	Pre-diabetic	Diabetic
Black males	100	22	78	30	51	19
Black females	94	19	81	48	44	9
Caucasian males	75	44	56	60	36	4
Caucasian females	94	45	55	82	17	1

The percentage of participants that were classified in the various blood pressure and blood glucose categories are given.

3.3 DNA ISOLATION, AMPLIFICATION AND PREPARATION FOR NEXT GENERATION SEQUENCING

Total genomic DNA was isolated from whole blood (EDTA-treated), which was thawed from frozen stock samples. The isolation of DNA was done using a Nucleospin® 96 blood kit (CAT# 740456.4) from Macherey-Nagel on a vacuum manifold, as per manufacturer instructions. For optimal recovered DNA yield and concentration, isolated DNA was eluted from the silica membrane in two steps, using 50 µL of heated (70 °C) ultra-pure, Type 1 water (Milli-Q®, Millipore) each time. DNA concentrations and quality were measured using a Nanodrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific).

Next, mtDNA was amplified in two overlapping fragments, using the AccuPrime™ Taq DNA polymerase High Fidelity kit from Invitrogen (CAT# 12346-094) as per manufacturer instructions. The 10X AccuPrime™ Buffer I was used, with 40 ng isolated DNA and 1 unit (0.2 µL) of AccuPrime™ High Fidelity enzyme, in a 50 µL reaction. For the primers used (Table 3.1), the optimal annealing temperature was 56.5 °C. For the first ten cycles, the duration of the extension step was kept constant at 10 min, after which it was sequentially extended by 10s after each next cycle (20X).

Table 3.3: Primers used for amplification of mtDNA

	Primer name	Primer sequence	Position on mtDNA	Size (bp)
Fragment A	SC-H-Forward	5'-ATCATACACAAACGCCTGAGC-3	13539-13559	9250
	SC-C-Reverse	5'-GGTAAGAGTCAGAAGCTTATG-3'	6200–6220	
Fragment B	SC-D-Forward	5'-AATACCCATCATAATCGGAGG-3'	6115–6135	7546
	SC-G-Reverse	5'-TTGACCTGTTAGGGTGAGAAGA-3'	13640–13660	

Two sets of primers were used to amplify mtDNA in two overlapping fragments. Here, the name, nucleotide sequence, position and size of each fragment is given.

To confirm that the correct mtDNA fragments were amplified, PCR products were separated on a 0.7% (w/v) agarose gel in Bionic™ Buffer from Sigma-Aldrich (CAT# B6185-10L), stained with 0.5 µg/ml gel ethidium bromide (EtBr), and visualised under UV-light using a Syngene G:BOX gel documentation system. A Thermo Scientific GeneRuler DNA Ladder mix (CAT# SM0334) was used to determine the relative size of DNA bands in the gel. Figure 3.1 is an example of such a gel.

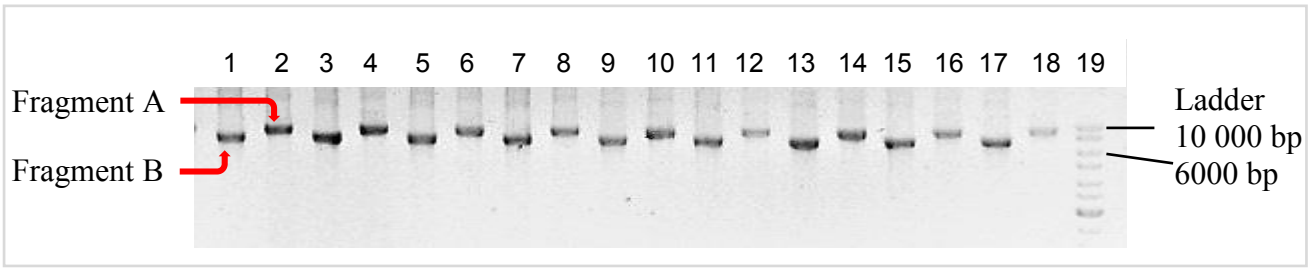


Figure 3.1: Image of agarose gel with mtDNA PCR products. Fragments A and B of several samples were added in alternating lanes (lanes 1-18). A DNA ladder was added in the lane 19.

Next, a PureLink™ Pro 96 PCR purification kit from Invitrogen (CAT# K3100-96A) was used to remove the remaining primers and other PCR reaction components from the PCR fragments. A vacuum manifold was again used together with the purification kit, per manufacturer instructions. To optimise recovered DNA yield, DNA was eluted in two steps, using 50-60 µL of low-Tris-EDTA (TE) buffer (pH 8.0) each time.

The purified DNA was then quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit from Invitrogen (CAT# P11496) and the fluorescence measured on a Synergy HT microplate reader. Finally, equimolar amounts of fragment A: fragment B (5.6:4.9) for each sample were combined, and the DNA concentration was adjusted to 10 ng/µL using low-TE buffer. A small amount of this fragment mix was then separated on a 0.7% (w/v) agarose gel in Bionic™ Buffer, to ensure by visual inspection that both fragments were indeed present and approximately in equimolar amounts (example in Figure 3.2). Samples where fragments where not visually of equimolar amount (see lane 11, Figure 3.2), were re-amplified or re-constituted. Although the samples amplified at different concentrations, as can be seen in this Figure 3.2, the concentrations were normalized for library preparation purposes. Samples were then stored at 4 °C until further use.

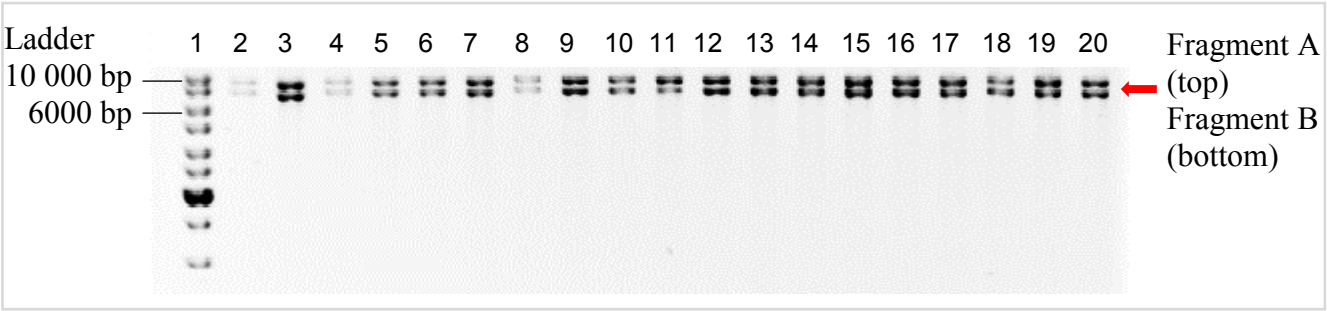


Figure 3.2: Image of agarose gel showing equimolar amounts of mtDNA fragments A and B. For each sample, fragment A and B were mixed and visually inspected on a 0.7% agarose gel. A DNA ladder was added in the first lane, so that the relative size of the bands could be estimated.

3.4 DNA LIBRARY BUILDING, ENRICHMENT AND TEMPLATE CONSTRUCTION FOR SEQUENCING

DNA libraries were constructed using the Ion Xpress™ Plus Library Kit (CAT# 4477598) and Ion Xpress™ Barcode Adapters 1-96 kit (CAT# 4474517) from Thermo Fisher Scientific, on a AB Library Builder™ System, as per manufacturer instructions, with some adjustments. Although 500 ng DNA was used for each reaction, only 2.5 µL of Ion Xpress™ P1 Adapter and Ion Xpress™ Barcode were added, after some optimisation runs. A custom shearing time was set at 9 minutes, resulting in DNA fragment of 150 to 300 bps long. A gel-based library size selection was done using an E-Gel® SizeSelect™ 2% Agarose kit from Invitrogen (CAT# G661002) on an E-Gel® electrophoresis system (Invitrogen).

Size-selected products were then quantified on a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific), using a Qubit® dsDNA HS assay kit from Thermo Fisher Scientific (CAT# Q32854), followed by amplification and purification as described in the Ion Xpress™ Plus and Ion Plus Library Preparation for the AB Library Builder™ System user guide. Equimolar amounts of amplified library products were then pooled using measurements from a combination of quantifying methods (Agilent® High Sensitivity DNA Kit (CAT# 5067-4626) and the Agilent 2100 Bioanalyzer® instrument and concentrations as measured on the Qubit® 2.0 Fluorometer). Libraries of 20 samples were pooled and diluted to ~13 pM.

Templates were then prepared using the Ion PGM™ Template OT2 Solutions 200 Kit (CAT# 4480974) and the Ion OneTouch™ 2 instrument, as per manufacturer instructions, to obtain template positive Ion Sphere™ Particles (ISPs). The quality of template positive ISPs were then assessed using the Ion Sphere™ Quality Control assay (CAT# 4468656) from Thermo Fisher Scientific, and the Qubit® 2.0 Fluorometer, as described in the Ion PGM™ Template OT2 200 kit user guide. The template positive ISPs were then enriched on the Ion OneTouch™ ES instrument, and stored at 4 °C for up to 3 days, before sequencing.

3.5 SEQUENCING OF MTDNA USING THE ION TORRENT PERSONAL GENOME MACHINE

Entire mtDNA sequences were obtained by synthesis method using the Ion PGM™ Sequencing 200 Kit v2 (CAT# 4482006), with the Ion 318™ Chip Kit v2 (CAT# 4484355), on the Ion Personal Genome Machine® (PGM™) system, all from Thermo Fisher Scientific. An Ion 318™ Chip was loaded manually after which the sequencing run was initiated. By sequencing only 20 samples per Ion 318™ Chip, the average coverage was estimated to be ~1800 times. In total, 21 templates were analysed. Figure 3.3 is a screenshot of part of a typical run summary that was obtained for each template. The parameters presented in such a run report can be used to assess the quality of a sequencing run.

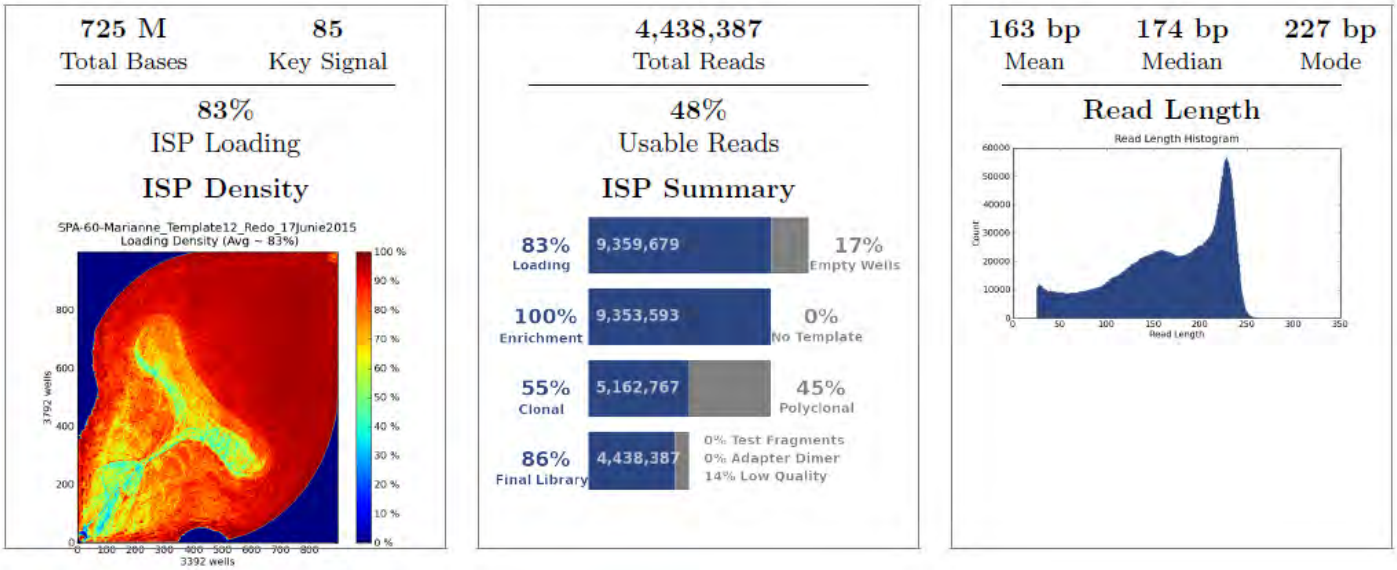


Figure 3.3: Screenshot of a run summary obtained for template 12. Several parameters pertaining to a sequencing run are given, including the percentage of chip that was loaded with ISPs, the percentage of ISPs that contained template/no template, and the percentage of ISPs that had more than one type of DNA fragment attached (polyclonal). In the last panel, the mean, median and mode for read lengths are given.

3.6 DATA ANALYSES

The Ion Torrent PGM™ produces binary versions of sequence alignment/map (BAM) and indexed BAM (BAI) files that can be used in several downstream processes. The Ion Torrent Suite 4.4 software automatically performs several steps such as filtering out reads of poor quality and trimming off primer and adapter sequences. Several steps were taken to ensure the quality of the data. Firstly, sequencing reads were trimmed to remove barcode and adapter sequences. Then a sliding window quality filter was applied with a quality score of 15. For this study, reads were aligned and variants called against the revised Cambridge Reference Sequence (rCRS) (Andrews et al., 1999) using the VariantCaller plug-in (version 5.0) unless specified otherwise. A list of candidate variants was first compiled, and these

variants were then evaluated using specific parameters (default settings), including read quality values, base quality values, variant frequency, strand bias, phase error, and coverage. The candidate variants not meeting the specific parameter requirements were filtered out. Furthermore, all variants used in the initial analyses had a quality score (Phred based scoring) above 50.3 (average for all variants =15054.3) and a minimum coverage of 6X (average for all variants = 1047X), while variants used in the final analyses (those with MutPred scores above 0.5), had quality scores above 62.2 (average=14224.5) and a minimum coverage of 7X (average = 1002.8X). The sequencing method has been validated using several other techniques. Previously, a subset (10%) of sequence data generated using the current Ion Torrent PGM™ sequencing protocol was compared with Miseq (Illumina) technology and sequence sections using pyrosequencing (QIAGEN). Although some differences existed in low frequency variants (not relevant here) a 100% match of consensus sequence data (relevant here) was obtained.

Aside from the specified websites and tools that are reported throughout this thesis, several online databases, tools and websites were used to perform unspecified procedures such as convert/transform file formats, inspect data, annotate variants and cross-reference information.

These included:

- *http://www.mitotool.org*
- *http://haplogrep.uibk.ac.at*
- *https://mtdna-server.uibk.ac.at/index.html*
- *http://www.phylotree.org*
- *http://www.mitomap.org;*
- *http://mamit-trna.u-strasbg.fr*
- *http://www.hugo-international.org/HUGO-Gene-Nomenclature*
- *http://www.ensembl.org/index.html*
- *https://www.ncbi.nlm.nih.gov/genbank*
- *http://mutpred.mutdb.org/*
- *https://galaxyproject.org/*

The specific data and statistical analysis methods used further to address each objective of this study, are described in the corresponding chapters that follow.

3.7 POWER CALCULATIONS

Power calculations included here were done using G*Power (Version 3.1.9.2). Considering mean MutPred adjusted load scores of 0.41 ± 0.28, for a total sample size of 82, this study has 90% power to detect a correlation with moderate effect size between MutPred adjusted loads and physiological parameters. When comparing mean MutPred adjusted loads between two phenotype classifications, this study had ~70% power to detect a medium effect size (d = 0.5) with a total sample size of 100.

Chapter 4: Evaluating the presence of disease-associated mtDNA variants as reported on MITOMAP, and mt-tRNA and mt-rRNA variant frequency in a hypertension cohort

4.1 INTRODUCTION

In Chapter 2, the various ways in which mitochondrial DNA (mtDNA) variants can be involved in disease were outlined. As discussed, clinically proven mtDNA point mutations are a recognised cause of primary mitochondrial disease. While strict criteria exist for an mtDNA variant to be classified as a "clinically proven" pathogenic mutation (see Chapter 2), many mtDNA variants have also been *associated* with disease (*www.mitomap.org*). In section 4.3.1 of this chapter, the presence in hypertension of these "disease-associated" variants, as reported in the mtDNA database MITOMAP (as of August, 2016), is investigated.

While variants in protein coding genes may cause specific structural or functional changes, variants in mt-tRNA genes might have a much wider impact, as these molecules are intricately involved in the transcription and translation of all mitochondrial protein genes (Florentz et al., 2003; Yarham et al.,

Selected glossary
Terms that could have ambiguous meanings or are new, are used throughout this thesis in the following way:
1) "mutation" here refers to any mtDNA variant that has been clinically proven, per set criteria as discussed in Chapter 2, to be pathogenic and cause disease. Mutations are thus "disease causing" variants.
2) "disease-associated" variants are mtDNA variants that have been associated with disease in some way, but do not necessarily at this point in time meet all the criteria set out to prove pathogenicity.
3) "MutPred scores" refer to pathogenicity scores which are assigned by the MutPred system (*mutpred.mutdb.org/about.html*) to any non-synonymous structural gene mtDNA variant, while "MutPred-scoring variant" refers to such variants. MutPred scores above 0.5 are considered an "actionable hypothesis" for pathogenicity, while scores above 0.75 are considered a "confident hypothesis".
4) "Yarham scores" refer to pathogenicity scores for mtDNA tRNA variants, as assigned by the system described by Yarham, et al., (2011). Variants are considered: neutral with a Yarham score below 7; possibly pathogenic with a score between 7 and 10; probably pathogenic with a score between 11 and 13, but with no function evidence such as single fibre, steady-state or cybrid analyses; and definitely pathogenic with a score above 10 and the inclusion of single fibre, steady-state or cybrid analyses evidence.

2010). Thus a severely disruptive mutation in, for instance, the mt-tRNA$^{Leu(UUR)}$ gene, will lead to decrease steady state levels of this mt-tRNA, decreased animoacylation and absence of the wobble base 5-taurinomethyl group modification, leading to reduced polypeptide synthesis where leucine is required, and inefficient enzyme complex assembly (Janssen et al., 2007). Likewise, mutations in mt-rRNA genes might affect the synthesis of all proteins expressed in the mitochondria. Unlike for mt-tRNAs, very few mt-rRNA mutations have been found (Elson et al., 2015). Two mutations in the 12S mt-rRNA, m.1555A>G and m.1494C>T, have been implicated in non-syndromic and aminoglycoside-induced deafness (Guan et al., 1996; Sylvester et al., 2004). It is thought that disease penetrance of these mutations is strongly modulated by other nuclear and environmental factors (Ballana et al., 2006; Guan et al., 2000; Raimundo et al., 2012; Zhao et al., 2005). In general, variants found in the stems rather than in the loops of the mt-tRNA molecule are more likely to be deleterious, as mutations here often disrupt Watson-Crick base pairs (Yarham et al., 2010). It is, however, also possible that the deleterious effect of such disruptions can be curtailed by a second-site nucleotide change which restores the Watson-Crick base pair (Kern & Kondrashov, 2004). It is also worth noting that some deleterious mutations are found in the loops of mt-tRNAs; these mutations might lead to codon:anticodon instability and negatively affect translation (Sanders et al., 2008).

While not all mt-tRNA or mt-rRNA variants are pathogenic, some variants might still infer some changes in mitochondrial function, and other downstream processes implicated in common complex disease, that is, they might be mildly deleterious. Mt-tRNA variants are good candidates for being mildly deleterious and involved in common complex disease, as they are more common in the recent "tips" of the phylogeny, rather than within the deeper branches (Elson et al., 2004; Kivisild et al., 2006). This indicates that selective pressure is exerted against these variants over time. Also, large families carrying deleterious mt-tRNA variants have been described, demonstrating that these variants are also transmitted to offspring (Qin et al., 2014). Therefore, it is possible that mt-tRNA or mt-rRNA variants might be more abundant in disease groups compared to control groups. In section 4.3.2 of this chapter, the presence of mt-tRNA and mt-rRNA variants in those with or without hypertension will be investigated.

4.2 METHODS

4.2.1 Cohort description, data generation and phenotyping

The cohort recruitment, ethics, sequencing data generation and variant calling was thoroughly described in Chapter 3. Variants were characterised using MITOMAP (*www.mitomap.org*). Haplogroups were assigned using Haplogrep 2.0 (*haplogrep.uibk.ac.at*) and MutPred scores were calculated for all non-synonymous variants using the MutPred program (*mutpred.mutdb.org/about.html*). Yarham scores were assigned as described by Yarham et al., (2011). Clinically proven mutations were identified using the criteria set out by DiMauro and Schon (2001), McFarland et al. (2004), Mitchell et al. (2006) and Yarham et al. (2011).

Large discrepancies in 24 hour ambulatory blood pressure monitoring (ABPM) measurements exist between Black South African and Caucasian South African SABPA participants, as well as between the two gender groups. For this reason, participants were divided into four gender/mtDNA background groups, and analysed separately. These were designated as follows: haplogroup L males and haplogroup L females (Black South African groups); haplogroup MN males and haplogroup MN females (Caucasian South African groups). For a detailed discussion on this, please refer to Chapter 5 (Venter et al., 2017). Normotensive and hypertensive classifications are based on 24 h ABPM measurements. A hypertensive classification was assigned to males with 24 h systolic ABPM ≥ 130 mmHg and/or 24 h diastolic ABPM ≥ 80 mmHg, and females with 24 h systolic ABPM ≥ 125 mmHg and/or 24 h diastolic ABPM ≥ 75 mmHg, based on findings reported in Hermida et al., (2013).

4.2.2 Statistical analyses

All statistical analyses were carried out using IBM SPSS Statistics (version 23) or Prism GraphPad software (version 6.05). For previously reported mtDNA variants, all variants were used so that heteroplasmy levels could be considered. To see whether these variants were more abundant in hypertensives than in normotensives, the number of times each variant appears in the cohort was counted. Fisher's exact tests were used to assess the differences in incidence of frequently occurring variants between the four gender/background groups. A Bonferroni correction for multiple testing was applied.

To assess the abundance of mt-tRNA and mt-rRNA variants, only consensus variants were used. First, the number of mt-tRNA and mt-rRNA variants that were very rare (defined as those with a GenBank frequency below 0.1%) were compared to those that were common in the population (defined as those with a GenBank frequency above 1%) using Fisher's exact tests. Next, mt-tRNA and mt-rRNA variants that were unique to each gender/background group (i.e. occurred only once in a group) were identified and counted. Fisher's exact tests were used to assess the significance of these frequencies.

4.3 RESULTS AND DISCUSSION

4.3.1 Frequency of previously reported disease-associated mtDNA variant in hypertension

In Table 4.1, variants present in the SABPA cohort that have been associated with disease, according to the MITOMAP database, are given, along with descriptive data for each variant. It is important to note that a number of these variants have weak evidence associating them with disease, but in this first instance an unbiased approach was taken by considering all variants reported as disease-associated. A similar approach to that was taken by Pereira et al. (2011), where all disease-associated mtDNA variants on the Online Mendelian Inheritance in Man (OMIM) database were used to investigate the role of selection in mtDNA evolution.

Table 4.1: Previously reported disease-associated mtDNA variants found in the SABPA cohort

Variant	Gene	AA change	MutPred or Yarham score	Conser-vation (%)	GB Frequency	Cohort frequency (counts)	Heteroplasmy (%)	MITOMAP disease report
9055G>A	*MT-ATP6*	A - T	0.492	87	1581 (4.93%)	9	>96	PD protective factor
9098T>C	*MT-ATP6*	I - T	0.685	76	47 (0.15%)	8	100	Predisposition to ARV mito disease
9185T>C	*MT-ATP6*	L - P	0.751	98	3 (0.01%)	1	5.9	LD / Ataxia syn / NARP-like disease
5911C>T	*MT-COI*	A - V	0.258	11	149 (0.46%)	27	>96	Prostate cancer
6253T>C	*MT-COI*	M - T	0.458	69	355 (1.11%)	6	100	Prostate cancer
6663A>G	*MT-COI*	I - V	0.558	96	132 (0.41%)	2	100	Prostate cancer
7080T>C	*MT-COI*	F - L	0.561	98	32 (0.10%)	3	one case with 7%	Prostate cancer
6261G>A	*MT-COI*	A - T	0.572	98	176 (0.55%)	2	>98	Prostate cancer / LHON
6267G>A	*MT-COI*	A - T	0.630	91	56 (0.17%)	2	>98	Prostate cancer
6489C>A	*MT-COI*	L - I	0.641	98	67 (0.21%)	1	98.7	Therapy-resistant epilepsy
6480G>A	*MT-COI*	V - I	0.664	89	121 (0.38%)	1	100	Prostate cancer
6150G>A	*MT-COI*	V - I	0.676	96	173 (0.54%)	6	>97	Prostate cancer
7859G>A	*MT-COII*	D - N	0.226	24	123 (0.38%)	1	100	Progressive encephalomyopathy
8021A>G	*MT-COII*	I - V	0.431	11	4 (0.01%)	2	100	Asthenozoospermia
9861T>C	*MT-COIII*	F - L	0.621	42	61 (0.19%)	2	one case with 5.8%	AD
9438G>A	*MT-COIII*	G - S	0.839	93	277 (0.86%)	3	>96	LHON
15693T>C	*MT-CYB*	M - T	0.229	76	396 (1.24%)	4	100	Possibly LVNC CM-associated
15497G>A	*MT-CYB*	G - S	0.477	89	158 (0.49%)	1	>96	EXIT / Obesity
15812G>A	*MT-CYB*	V - M	0.518	24	212 (0.66%)	8	>97	LHON
15257G>A	*MT-CYB*	D - N	0.785	96	405 (1.26%)	8	>99	LHON
15043G>A	*MT-CYB*	G - G	-	98	7301 (22.77%)	17	>96	DEAF associated
3310C>T	*MT-ND1*	P - S	0.434	13	9 (0.03%)	1	100	Diabetes / HCM
3316G>A	*MT-ND1*	A - T	0.463	4	303 (0.95%)	3	100	Diabetes / LHON / PEO
3796A>G	*MT-ND1*	T - A	0.540	71	174 (0.54%)	2	>97	Adult-onset dystonia
3308T>C	*MT-ND1*	M - T	0.577	84	306 (0.95%)	1	100	MELAS / DEAF enhancer / HT / LVNC
3337G>A	*MT-ND1*	V - M	0.600	24	52 (0.16%)	3	100	CM
4216T>C	*MT-ND1*	Y - H	0.611	24	3156 (9.84%)	40	one case with 4.5%	LHON / IR / adaptive
3736G>A	*MT-ND1*	V - I	0.661	93	59 (0.18%)	2	>99	LHON
3644T>C	*MT-ND1*	V - A	0.676	100	154 (0.48%)	1	98.9	BD-associated

Table 4.1 continued...

Variant	Gene	AA change	MutPred or Yarham score	Conser-vation (%)	GB Frequency	Cohort frequency (counts)	Heteroplasmy (%)	MITOMAP disease report
3396T>C	*MT-ND1*	Y - Y	-	93	266 (0.83%)	1	100	NSHL / MIDD
5460G>A	*MT-ND2*	A - T	0.505	4	2003 (6.25%)	62	>96	AD / PD
4917A>G	*MT-ND2*	N - D	0.628	91	1528 (4.77%)	25	>90	LHON / IR / AMD / NRTI-PN
10398A>G	*MT-ND3*	T - A	0.170	51	14003 (43.68%)	250	one case with 4.4%	Various
10086A>G	*MT-ND3*	N - D	0.196	87	175 (0.55%)	11	>96	Hypertensive end-stage renal disease
10237T>C	*MT-ND3*	I - T	0.551	100	50 (0.16%)	1	97.7	LHON
11084A>G	*MT-ND4*	T - A	0.482	87	159 (0.50%)	1	>96	AD, PD; MELAS
11778G>A	*MT-ND4*	R - H	0.852	100	112 (0.35%)	1	16.5	LHON / Progressive dystonia
11467A>G	*MT-ND4*	L - L	-	100	4213 (13.14%)	44	>93	Altered brain pH
10652T>C	*MT-ND4L*	I - I	-	73	51 (0.16%)	1	4.9	BD / MIDD-associated
12372G>A	*MT-ND5*	L - L	-	80	4519 (14.10%)	45	>96	Altered brain pH
12634A>G	*MT-ND5*	I - V	0.381	98	91 (0.28%)	1	100	Thyroid cancer cell line
13708G>A	*MT-ND5*	A - T	0.409	33	2185 (6.82%)	21	>96	LHON / Increased MS risk
13637A>G	*MT-ND5*	Q - R	0.610	62	287 (0.90%)	3	>95	Possible LHON factor
669T>C	*MT-RNR1*	rRNA	-	58	66 (0.21%)	6	>95	DEAF
721T>C	*MT-RNR1*	rRNA	-	4	58 (0.18%)	6	>99	Possibly LVNC-associated
921T>C	*MT-RNR1*	rRNA	-	67	221 (0.69%)	8	100	Possibly LVNC-associated
961T>C	*MT-RNR1*	rRNA	-	22	317 (0.99%)	25	two cases < 5%; one case with 56%; one case with 86%	DEAF, possibly LVNC-associated
961T>G	*MT-RNR1*	rRNA	-	22	123 (0.38%)	1	99.4	Possibly DEAF-associated
1310C>T	*MT-RNR1*	rRNA	-	29	30 (0.09%)	1	98.2	DEAF-associated
2352T>C	*MT-RNR2*	rRNA	-	9	839 (2.62%)	19	one case with 9.6%	Possibly LVNC-associated
2755A>G	*MT-RNR2*	rRNA	-	56	184 (0.57%)	7	>98	Possibly LVNC-associated
3010G>A	*MT-RNR2*	rRNA	-	20	5046 (15.74%)	34	one case with 5.1%	Cyclic vomiting syndrome/migraine
3196G>A	*MT-RNR2*	rRNA	-	22	8 (0.02%)	1	8.4	ADPD
5655T>C	*MT-TA*	tRNA	5	0	276 (0.86%)	3	100	DEAF enhancer
5814T>C	*MT-TC*	tRNA	14	16	109 (0.34%)	1	100	Mitochondrial encephalopathy
7543A>G	*MT-TD*	tRNA	9	84	8 (0.02%)	1	98.7	MEPR

Table 4.1 continued...

Variant	Gene	AA change	MutPred or Yarham score	Conser-vation (%)	GenBank Frequency	Cohort frequency (counts)	Heteroplasmy (%)	MITOMAP disease report
14687A>G	*MT-TE*	tRNA	13	2	211 (0.66%)	3	100	Mito myopathy w respiratory failure
3277G>A	*MT-TL1*	tRNA	0	11	19 (0.06%)	4	100	Poss. HT factor
4336T>C	*MT-TQ*	tRNA	8	36	334 (1.04%)	1	100	ADPD / Hearing loss & migraine
4395A>G	*MT-TQ*	tRNA	0	7	19 (0.06%)	1	100	Poss. HT factor
4454T>C	*MT-TM*	tRNA	0	56	185 (0.58%)	1	100	Poss. mito dysfunction / HT
8343A>G	*MT-TK*	tRNA	0	18	26 (0.08%)	3	one case with 5.1%	Possible PD risk factor
10454T>C	*MT-TR*	tRNA	1	36	118 (0.37%)	1	98	DEAF helper mut.
12192G>A	*MT-TH*	tRNA	4	7	72 (0.22%)	8	>98	MICM
12236G>A	*MT-TS2*	tRNA	0	71	284 (0.89%)	3	one case with 5.3%	DEAF
12308A>G	*MT-TL2*	tRNA	0	96	4193 (13.08%)	44	>87	CPEO / Stroke / CM / other
15924A>G	*MT-TT*	tRNA	2	71	1288 (4.02%)	11	one case with 10.1%	LIMM
15928G>A	*MT-TT*	tRNA	0	49	1547 (4.83%)	24	>90	Various
15942T>C	*MT-TT*	tRNA	0	16	237 (0.74%)	13	one case with 10.9%	Possibly LVNC-associated
15951A>G	*MT-TT*	tRNA	6	76	308 (0.96%)	4	100	LHON modulator

In this table, variants that have previously been reported in association with disease according to MITOMAP (*www.mitomap.org*) that are present in the SABPA cohort are listed. For each variant, the locus and amino acid (AA) change are given, as well as the MutPred score (*mutpred.mutdb.org/about.html*) for non-synonymous variants, or Yarham score (as described in Yarham et al., 2011) for mt-tRNA variants. Conservation is across 45 species. GenBank frequency was derived by MITOMAP from 32 069 complete sequences available on GenBank as of August, 2016. Cohort frequency refers to the number of times a variant appears within the SABPA cohort. Heteroplasmy levels of these variants within SABPA participants are also given; where heteroplasmy levels of single cases are specified, all other cases were homoplasmic. Transversions are indicated in italics. Clinically proven mutations are indicated in boldface. AD: Alzheimer's disease; ADPD: Alzheimer's disease and Parkinson's Disease; AMD: age-related macular degeneration; BD: bi-polar disorder; CM: cardiomyopathy; CPEO: chronic progressive external ophthalmoplegia; DEAF: maternally inherited deafness or aminoglycoside-induced deafness; EXIT: exercise intolerance; HT: hypertension; IR: insulin resistance; LHON: Leber's hereditary optic neuropathy; LIMM: lethal infantile mitochondrial myopathy; LVNC: left ventricular non-compaction; MELAS: mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes; MEPR: myoclonic seizures, developmental delay, and severe behavioural problems; MICM: maternally inherited cardiomyopathy; MIDD: maternally inherited deafness and diabetes; MM: mitochondrial myopathy; NARP: neurogenic muscle weakness, ataxia, and retinitis pigmentosa; NRTI-PN: antiretroviral therapy-associated peripheral neuropathy; SNHL: sensory neural hearing loss.

Several disease-associated variants occurred multiple times in the SABPA cohort, though most of these were evenly distributed among the hypertensive and normotensive groups. All haplogroup L participants had at least one disease-associated variant, with the exception of one haplogroup L female participant (hypertensive), who had none. In haplogroup MN, only five hypertensive and four normotensive males, as well as four hypertensive and ten normotensive females had no disease-associated variants. All other haplogroup MN participants had at least one disease-associated variant. In total, disease-associated variants occurred 166 times in hypertensive haplogroup L males ($N = 78$), and 39 times in normotensive haplogroup L males ($N = 22$). Although this appears to be a significant difference, it is important to note the large difference in the number participants in the hypertensive and normotensive haplogroup L male groups. Expressed as the number of variants per person, these frequencies are 2.1 for hypertensives and 1.8 for normotensives, and were not statistically significant (Fisher's $P = 0.35$). The same was true for all other groups: disease-associated variants occurred 161 times in hypertensive haplogroup L females ($N = 76$) and 36 times in normotensive haplogroup L females ($N = 18$), which was 2.12 and 2 variants per person for hypertensives and normotensives respectively; variants occurred 99 times in hypertensive haplogroup MN males ($N = 44$) and 101 times in normotensive haplogroup MN males ($N = 46$), which was 2.3 and 3.1 variants per person for hypertensives and normotensives respectively; lastly, variants occurred 119 times in hypertensive haplogroup MN females ($N = 52$) and 95 times in normotensive haplogroup MN females ($N = 42$), which was 2.3 variants per person for both blood pressure groups. None of these frequency distributions were significant (Fisher's $P > 0.05$).

When looked at individually, most of these variants were evenly distributed among hypertensive and normotensive groups. However, some variants appeared to segregate better with phenotype groups. In Table 4.2, four disease-associated variants with the most significant differences in distribution among hypertensive and normotensive participants are listed. For each gender/background group, the percentage of participants in which a specific variant occurred is given.

Table 4.2: Cohort frequency of four frequently occurring previously reported disease-associated mtDNA variants

Group	Mean group systolic 24h ABPM ± SD (mmHg)	Blood pressure status	Percentage of participants in which the variant is present (Yes) or absent (No) (%)							
			961T>C		5460G>A		5911C>T		10398A>G	
			Yes	No	Yes	No	Yes	No	Yes	No
Haplogroup L males	143.2 ± 13	Hypertensive (*N* = 78)	18*	82	32	68	18*	82	97	3
	118.2 ± 5	Normotensive (*N* = 22)	0	100	14	86	0	100	100	0
Haplogroup L females	132.6 ± 14	Hypertensive (*N* = 76)	12	88	26	74	11	89	97	3
	111.6 ± 4	Normotensive (*N* = 18)	0	100	22	78	0	100	100	0
Haplogroup MN males	133.9 ± 10	Hypertensive (*N* = 42)	0	100	2	98	0	100	17	83
	120.7 ± 5	Normotensive (*N* = 33)	0	100	3	97	0	100	24	76
Haplogroup MN females	127.8 ± 13	Hypertensive (*N* = 52)	0	100	6	94	0	100	21	79
	112.0 ± 4	Normotensive (*N* = 42)	0	100	2	98	0	100	31	69

In this table, the percentage of participants in a gender/background group, in which a specific variant occurred is given. Distributions that were significant at the 0.05 level (before correction of multiple testing) are marked with an asterisk. ABPM: ambulatory blood pressure measurements.

Although several more hypertensive haplogroup L participants (both males and females) had variant m.5460G>A than their normotensive counterparts, these frequencies were not statistically significant (Fisher's $P > 0.1$). The m.10398A>G variant has a history of being reported in association with complex disease, ranging from being both a protective factor against (van der Walt et al., 2003) and a risk factor for (Otaegui et al., 2004) Parkinson's disease, as well as a risk factor for metabolic syndrome and breast cancer (Canter et al., 2005), which has been directly disputed (Setiawan et al., 2008). m.10398A>G currently has a high GenBank frequency of 44% (present in 14003 sequences on GenBank), and a very low MutPred pathogenicity score of 0.17, which makes it very unlikely that this variant, on its own, is deleterious. Both m.961T>C and m.5911C>T were only present in haplogroup L participants, and only in those with hypertension. This was statistically significant in haplogroup L males (Fisher's $P = 0.03$) but not females, but significance was lost after a Bonferroni correction. Both of these variants are listed on Phylotree as haplogroup-defining (m.5911C>T defines L0a1b, while m.961T>C defines a sub-haplogroup of L0a1b, namely L0a1b1a1). Interestingly, all participants in this cohort that had both these variants, and were assigned haplogroups L0a1b ($N = 1$) or L0a1b1a1 ($N = 21$), were also classified as hypertensive. One participant who only had the m.961T>C variant and was assigned haplogroup L0d2a1, was also hypertensive. It should be noted however that: i) the number of haplogroup L participants classified as hypertensive is unusually high, given that participants were not selected on the grounds of phenotype (i.e. hypertensive participants were not sought out during recruitment). This is probably an indication of

population bias, as clinically relevant cut-off values of hypertension classifications are based on studies mostly done on Caucasians. It is likely that these cut-off values should be adjusted for use in different population groups; ii) while it is tempting to associate hypertension with haplogroup L01ab and its subgroups, this study is not powered well enough to do a haplogroup association (Samuels et al., 2006). Thus, while this result is interesting to consider, it is by no means a confident indication of such an association. Even though these four variants have previously been associated with disease, as common haplogroup markers occurring multiple times in the phylogeny and having mostly low MutPred scores, it is unlikely that they are disease- causing variants with direct effects on structure or function of mtDNA proteins. However, as mentioned in Chapter 2, haplogroup context could in some way influence risk for disease progression or outcome (Achilli et al., 2011; Chinnery et al., 2010; Hudson et al., 2007; Ji et al., 2008).

As negative selection removes pathogenic mtDNA variants over time, rare variants are more likely to be deleterious (Elson et al., 2004; Pereira et al., 2011). In this cohort, only nine very rare (GenBank frequency less than 0.1%) disease-associated variants were found, and these variants were also rare among participants in this cohort.

Three clinically-proven mutations associated with mitochondrial disease phenotypes were present in the SABPA cohort (indicated in bold in Table 4.1). The first, found in *MT-ND4*, is one of the three most common Leber's hereditary optic neuropathy (LHON) mutations that account for 90-95% of LHON cases (Yu-Wai-Man & Chinnery, 2000 [Updated 2016]). LHON mutations cause a very tissue-specific phenotype (bilateral visual failure), and are usually homoplasmic. Disease penetrance usually occurs only at heteroplasmy levels above 60%, as measured in leucocytes, and is greatly influenced by other factors such as age, gender, smoking and haplogroup background (Chapter 2), but the phenotype can also be completely absent even at homoplasmic mutation levels. As this mutation was detected in blood at a heteroplasmy level of 16.5% in one 56-year-old male participant here, it is unlikely to have any biologically relevant effect in the participant now of even later in life.

The next two clinically-proven mutations were m.5814T>C in *MT-TC*, found in one haplogroup L female, and m.14687A>G in *MT-TE*, found in three haplogroup MN females. The m.5814T>C mutation is implicated in mitochondrial encephalopathy (Sternberg et al., 2001), while m.14687A>G is implicated in mitochondrial myopathy with respiratory failure (Bruno et al., 2003). Yarham scores of 14 and 13 for m.5814T>C and m.14687A>G respectively have previously been assigned (Supplementary information of Yarham et al., 2011). Of the four participants in whom these mutations occurred, three were borderline hypertensive, while one participant (with m.14687A>G) had significantly high measurements of 24 h ABPM and reactive oxygen species (ROS). One participant (with m.5814T>C) had high inflammation (C-reactive

protein) and albuminuria measurements. Yarham scores take into account several factors such as patient segregation, conservation across species, heteroplasmy levels, and biochemical, histochemical and function evidence to assign pathogenicity scores. Although both these mutations meet several criteria for pathogenicity, both mutations are also listed on Phylotree (Built 17), as haplogroup defining variants, each for the corresponding haplogroups that were assigned to the participants (L2 or T2). Also, both mutations were present at homoplasmic levels in these participants, which would be unusual for pathogenic mutations (LHON mutations being an exception to this rule). While the presence of a variant on Phylotree alone is not enough to disqualify it as disease-causing, it does call for careful inspection of the evidence for pathogenicity previously put forth. The functional evidence for pathogenicity for both mutations was reported more than a decade ago (Bruno et al., 2003; Sternberg et al., 2001). It is very likely that at the time, these mutations were considered to be rare, and not yet classified as population variants. On closer inspection, m.5814T>C did not have the single fibre analysis evidence as previously thought (Sternberg et al., 2001), which reduced its Yarham score from 14 to 11, placing it in the "probably pathogenic" category. Although single fibre analysis evidence for m.14687A>G is presented by Bruno et al. (2003), the difference between negative and positive COX stained fibres is not very large (statistically significant at the 0.01 level). Also, in this particular study, only the mt-tRNA genes were sequenced. It is therefore possible that the pathogenicity the authors detected could have been caused by another variant present elsewhere on the patient mtDNA (Bruno et al., 2003). It is likely that this mutation should also be reclassified as "probably pathogenic". Since it is unlikely that these variants are pathogenic, it is also unlikely that the participants carrying these variants will present with the associated disease phenotypes now or later in life. These cases are good illustrations of why it is important to re-evaluate the pathogenic status of variants/mutations on a regular basis, as it could have substantial implications in the clinical field of primary mitochondrial disease diagnosis and genetic counselling.

4.3.2 *Mitochondrial tRNA and rRNA variant frequency in hypertension*

mtDNA is subject to selection against deleterious variants, and as such, rare variants are more likely to be involved in disease (Elson et al., 2004). To assess the abundance of mt-tRNA and mt-rRNA variants in hypertensive vs normotensive groups, the frequency of variants that were very rare in the population (GenBank frequency below 0.1%) was compared to that of common variants (GenBank frequency above 1%). These are listed in Table 4.4. When considering the data in this table, it should be noted that for haplogroup L groups, there are almost four times more participants classified as hypertensive than normotensive. Subsequently, even though no very rare mt-tRNA variants were found in normotensive haplogroup L participants, while six such variants were found in hypertensive haplogroup L males and two in hypertensive haplogroup L females, this was not a significant different. Hypertensive haplogroup L participants were also no more likely to have very rare mt-rRNA variants, than their normotensive counterparts. Similarly, hypertensive haplogroup MN participants were no more likely to harbour very rare mt-tRNA or mt-rRNA variants than their normotensive counterparts. The differences in distribution

between very rare and common variants among hypertensive and normotensive groups was also not statistically significant.

Next, unique variants for each gender/background group (i.e. variants only present once in each group) were identified. Figure 4.1 is a graphical representation of these frequencies for mt-tRNAs (panel A) and mt-rRNAs (panel B). As with the disease-associated variants in the previous section, unique mt-tRNA and mt-rRNA variants occurred more times in hypertensive haplogroup L male and female groups than in normotensive groups. In haplogroup MN groups the distribution among hypertensive and normotensive groups was more even. It is again important to take note of the number of haplogroup L participants that are classified as hypertensive. It is very likely that this unequal distribution of haplogroup L participants among blood pressure classification groups, contributes to the distribution of mt-tRNA and mt-rRNA variants seen in Figure 4.1. Indeed, none of these differences were of statistical significance (Fisher's $P > 0.1$). It therefore seems that the number of unique mt-tRNA and mt-rRNA variants present in each group does not significantly contribute to the development of hypertension.

Table 4.4: Number of rare and common mt-RNA variants per group

mt-RNA	GenBank frequency (%)	Male			Female		
		Hyper-tensive	Normo-tensive	Fisher's *P*	Hyper-tensive	Normo-tensive	Fisher's *P*
mt-tRNA variants	*Haplogroup L*	(*N*=78)	(*N*=22)		(*N*=76)	(*N*=18)	
	< 0.1 %	6	0	0.11	2	0	0.49
	> 1 %	7	6		7	6	
	Haplogroup MN						
	< 0.1 %	2	1	1	1	1	1
	> 1 %	7	5		5	7	
mt-rRNA variants	*Haplogroup L*	(*N*=42)	(*N*=33)		(*N*=52)	(*N*=42)	
	< 0.1 %	4	2	0.67	4	4	1
	> 1 %	18	16		18	14	
	Haplogroup MN						
	< 0.1 %	4	2	0.64	4	5	1
	> 1 %	9	11		10	12	

The number of very rare (GenBank frequency < 0.1%) and common (GenBank frequency > 1%) mt-tRNA and mt-rRNA variants are presented for each gender/background group. Fisher's exact tests were used to test the significance of the distribution of these variants among the gender/background groups. The number of participants (*N*) for each group are also indicated.

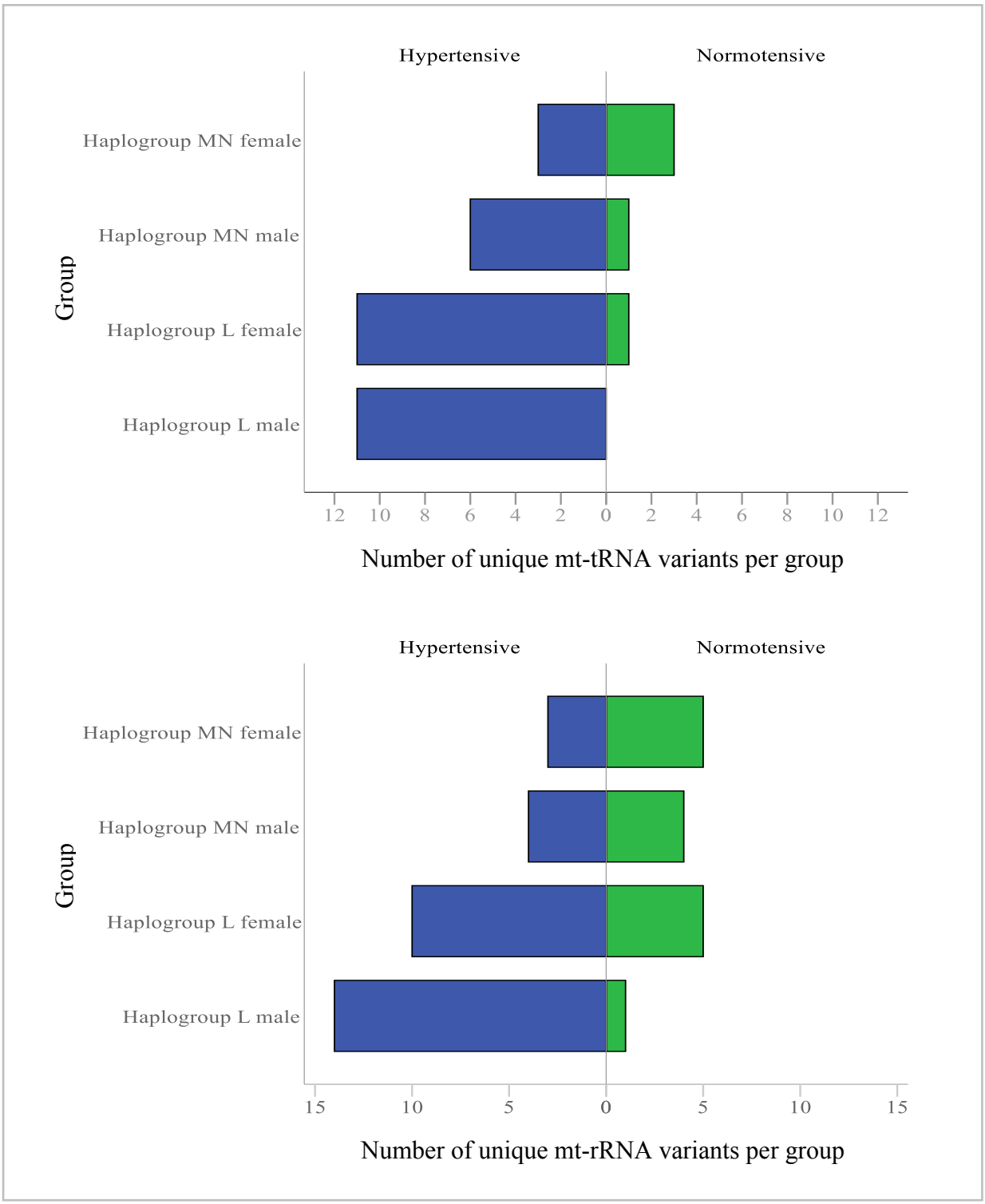


Figure 4.1: Distribution of mt-tRNA (A) and mt-rRNA (B) variants among gender/background groups. In this figure, the number of mt-tRNA or mt-rRNA variants that are unique for each gender/background group is given. Unique variants are those that occur only once in each gender/background group. The number of participants (*N*) for each group were as follows: hypertensive haplogroup L males: 78; normotensive haplogroup L males: 22; hypertensive haplogroup L females: 76; normotensive haplogroup L females: 18; hypertensive haplogroup MN males: 42; normotensive haplogroup MN males: 33; hypertensive haplogroup MN females: 52; normotensive haplogroup MN females: 42.

4.4 CONCLUDING REMARKS

While clinically proven mtDNA mutations play a key role in primary mitochondrial disease, numerous other rare and common mtDNA variants have been associated with disease. Although several of these disease-associated variants were present in the SABPA cohort, none were associated in a statistically significant way with hypertension. It is likely that many of these variants are not truly associated with disease, as several issues exist for the current approaches used when mtDNA variants are investigated in the context of common/complex disease. The loss of significance after a correction for multiple testing in the analyses described above also demonstrates the high statistical burden imposed by the use of multiple individual tests, as is done in haplogroup association studies. In the next chapter, a new approach is described and applied, which might help to address some of these points.

A range of disease phenotypes are caused by mt-tRNA and mt-rRNA mutations. Since mt-tRNA variants are more common in the recent "tips" of the phylogeny, they are likely to be mildly deleterious, and could possibly be more abundant in disease groups than control groups. However, the number of very rare (GenBank frequency below 0.1%) or unique (occurring only once per group) mt-tRNA and mt-rRNA variants did not significantly differ between hypertensive and normotensive groups, and likely does not contribute to disease presentation. Frequency analyses throughout this study were influenced by the possible bias introduced by phenotypical reference value thresholds that might not be uniformly applicable to all population groups. Not only does this highlight the need for more phenotypical and genetic studies to be done on these populations, but also the importance of designing such studies with appropriate case-and-control groups in mind.

Chapter 5: Evaluating the role of MutPred adjusted loads in hypertension and hyperglycaemia

This chapter consists of a peer-reviewed paper.

Published paper: **Using MutPred derived mtDNA load scores to evaluate mtDNA variation in hypertension and diabetes in a two-population cohort: The SABPA study**

Marianne Venter, Leone Malan, Etresia van Dyk, Joanna L. Elson, Francois H. van der Westhuizen

Published in *Journal of Genetics and Genomics*. 44 (2017): 139-149

http://dx.doi.org/10.1016/j.jgg.2016.12.003

Journal of Genetics and Genomics 44 (2017) 139–149

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

Original research

Using MutPred derived mtDNA load scores to evaluate mtDNA variation in hypertension and diabetes in a two-population cohort: The SABPA study

Marianne Venter [a], Leone Malan [b], Etresia van Dyk [a], Joanna L. Elson [a, c, *, 1], Francois H. van der Westhuizen [a, 1]

[a] Human Metabolomics, North-West University, Potchefstroom 2531, South Africa
[b] Hypertension in Africa Research Team (HART), North-West University, Potchefstroom 2531, South Africa
[c] Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne NE1 3BZ, United Kingdom

ARTICLE INFO

Article history:
Received 24 August 2016
Received in revised form
11 December 2016
Accepted 22 December 2016
Available online 26 December 2016

Keywords:
Mitochondrial DNA
MutPred
Mutational load
Hypertension
Diabetes
African
SABPA

ABSTRACT

Mitochondrial DNA (mtDNA) variation has been implicated in many common complex diseases, but inconsistent and contradicting results are common. Here we introduce a novel mutational load hypothesis, which also considers the collective effect of mainly rare variants, utilising the MutPred Program. We apply this new methodology to investigate the possible role of mtDNA in two cardiovascular disease (CVD) phenotypes (hypertension and hyperglycaemia), within a two-population cohort ($n = 363$; mean age 45 ± 9 yrs). Very few studies have looked at African mtDNA variation in the context of complex disease, and none using complete sequence data in a well-phenotyped cohort. As such, our study will also extend our knowledge of African mtDNA variation, with complete sequences of Southern Africans being especially under-represented. The cohort showed prevalence rates for hypertension (58.6%) and prediabetes (44.8%). We could not identify a statistically significant role for mtDNA variation in association with hypertension or hyperglycaemia in our cohort. However, we are of the opinion that the method described will find wide application in the field, being especially useful for cohorts from multiple locations or with a variety of mtDNA lineages, where the traditional haplogroup association method has been particularly likely to generate spurious results in the context of association with common complex disease.

1. Introduction

Cardiovascular disease (CVD) is an umbrella term that encompasses several distinctive disease phenotypes incisive of myocardial infarction, stroke, congenital heart disease and risk factors such as hypertension, and diabetes type 2 (Mensah, 2013). Discrepancies between different population groups exist in the onset, development and outcome of CVDs (Okin et al., 2011; Moran et al., 2013; Owolabi et al., 2014). Although environmental and lifestyle factors play a role in the risk of developing CVD, genetic factors are likely to account for some of the observed discrepancies in CVD onset, and particularly progression/outcome of disease (Achilli et al., 2011). Nuclear DNA risk factors have not been able to account for all the observed inconsistencies among different population groups (Kaufman et al., 2015). Thus, other sources of variation, such as epigenetics and mitochondrial DNA (mtDNA) variation might account for some of the missing heritability in CVDs.

mtDNA mutations are a common cause of inherited disease (Gorman et al., 2015). Both diabetes and cardiomyopathy are frequent symptoms in primary mitochondrial disease (Taylor et al., 2003; Yarham et al., 2010), especially in patients with the mtDNA m.3243A>G mutation (Hollingsworth et al., 2012). While clinically manifesting mtDNA mutations are a recognised cause of human

Abbreviation: 24 h ABPM, 24 h ambulatory blood pressure monitoring; CVD, cardiovascular disease; HbA1c, glycated haemoglobin; mtDNA, mitochondrial deoxyribonucleic acid; SABPA, Sympathetic Activity and Ambulatory Blood Pressure in Africans.

* Corresponding author.
E-mail address: j.l.elson@ncl.ac.uk (J.L. Elson).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jgg.2016.12.003

disease, many studies have suggested a role for common mtDNA variants (Hernstadt and Howell, 2004; Wallace, 2010) and the combined effect of rare population variants (Elson et al., 2006) in common complex disease. For CVDs specifically, a more recent Framingham Heart sub-study showed significant associations between several population as well as rare variants, and variation in blood pressure and fasting blood glucose levels (Liu et al., 2012). Cardena et al. (2014, 2016) found the hypertensive phenotype to be more prevalent among heart failure patients with an African mitochondrial haplogroup than among those with European haplogroups, in an admixed Brazilian cohort.

Human mtDNA codes for 13 essential polypeptide components of the mitochondrial oxidative phosphorylation (OXPHOS) system. mtDNA undergoes strict maternal inheritance, resulting in the absence of bi-parental recombination (Elson et al., 2001), and has a high mutation rate (Tuppen et al., 2010). As such, the evolution of mtDNA is characterised by the emergence of distinct lineages (or haplogroups) (Hernstadt et al., 2002). This results in high levels of mtDNA variation at the population level despite its rather small size, which is also illustrated by the large number of sub-haplogroups (van Oven and Kayser, 2009). Africa (haplogroups L0–L6) has the highest levels of nuclear and mitochondrial genetic diversity (Salas et al., 2002); this diversity was reduced by population bottlenecks in groups migrating from Africa (Manica et al., 2007). However, variation in African populations, especially in relation to disease, is still under studied when compared to variation in super clade N (Cavadas et al., 2015; Gurdasani et al., 2015). Only about 12% of mtDNA sequences on GenBank are of African lineages. In terms of mtDNA ancestry, the African haplogroup L3 incorporates super clades M and N, which encompass all the European and Asian haplogroups (Hernstadt et al., 2002; van Oven and Kayser, 2009; Rosa and Brehem, 2011).

Given the unique inheritance pattern of mtDNA, it is worth considering the hypothesis that might link mtDNA population variation to common complex disease. Firstly, studies showing significant associations of several disease phenotypes with specific mtDNA haplogroups have been published, suggesting that one or more common population variants may modify risk or outcome of disease. Secondly, the high mutation rate of mtDNA frequently results in the same variant being present more than once on the phylogeny (Hernstadt et al., 2002). Showing that such a variant modifies risk or alters the course of disease (Yu et al., 2008) in two different haplogroups or global populations, would be excellent evidence of its role in disease. The third possibility is that rare mtDNA mutations might have an effect, either at the individual level with there being synergist effects of multiple rare variants in patients, or rare variants might just be seen more frequently in the patient group.

The first hypothesis mentioned above, known as the haplogroup association hypothesis, has to date been the classic approach when considering the role of mtDNA variation in common disease. These studies have been controversial due to their low repeatability; often the association is not detected in a second cohort, or an association with a different haplogroup might be uncovered (Salas and Elson, 2015). Some studies have taken a two-cohort approach in an attempt to address these problems (Elson et al., 2006; Chinnery et al., 2010). Even so, this approach has proved unsuitable for many studies with cohorts that do not have the large numbers required for well-powered haplogroup association studies (Samuels et al., 2006).

In this study, we took an alternative approach, using MutPred pathogenicity scores to derive "mutational loads". This is a new version of the "mutational load" hypothesis proposed by Elson et al. (2006), which looked at the frequency of mildly deleterious (rare) variants in patient and control groups, to test the possibility of a cumulative effect of these variants. After calculating "MutPred mutational loads" by summing the MutPred scores generated for each of the non-synonymous variants on an individual's mtDNA, we adjusted these for the position of the sequence in the phylogeny, ultimately calculating "MutPred adjusted loads". Many different tools for predicting pathogenicity of mtDNA variants exist, but in a comparative study by Thusberg et al. (2011), MutPred and SNPs&GO outperformed all other methods, which included PolyPhen2, SIFT and SNAP. The MutPred program has also been widely validated in the context of mtDNA variation (Pereira et al., 2011). The MutPred algorithm incorporates elements of the SIFT algorithm, and assigns a pathogenicity score between 0 and 1, with zero being a benign substitution. A pathogenic score above 0.5 can be considered an "actionable hypothesis", while a score above 0.75 can be considered a "confident hypothesis" (Li et al., 2009). As there is selection against mildly deleterious variants at the population level (Elson et al., 2004; Soares et al., 2013), variants with MutPred scores above the "actionable hypothesis" threshold (0.5) are less likely to define major haplogroups; rather, they are more likely to be rare and seen on very recent branches (twigs) of the phylogeny (Pereira et al., 2011). Because the number of low-scoring, low-impact, common variants differs greatly among different population groups, their inclusion in the calculation of mutational loads could introduce noise that is unlikely to have phenotypic impact. By excluding variants with a pathogenicity score below 0.5, and thus most common population variants, before calculating MutPred adjusted loads, we aim to highlight the impact of rare variants while reducing the effect of population stratification.

Although mtDNA variation involvement in CVDs has previously been investigated (Chinnery et al., 2010), these studies have been predominantly in Caucasian European populations. Very few such studies have been published on African populations or those of African descent; most were on small sized cohorts, and often focused on specific previously reported variants (Khogali et al., 2001; Robinson et al., 2004; Ameh et al., 2011). To assume that findings from studies in European populations can easily be extrapolated to be less investigated, genetically diverse African population groups would be short sighted (van der Westhuizen et al., 2015), as even clinically proven disease causing mutations as well as the underlying spectrum of mutations are known to have differing impacts in Africans (van der Walt et al., 2012; van der Westhuizen et al., 2015).

In this study, we used the two-population Sympathetic Activity and Ambulatory Blood Pressure in Africans (SABPA) cohort (Malan et al., 2015) (Table 1). This is a South African cohort which consisted of 409 Black and Caucasian South African participants from the same geographical area (Rosa and Brehem, 2011; Salas and Elson, 2015). Although the participants were matched for gender, age and socio-economic status, a significantly higher percentage of Black participants, compared to Caucasians participants, with optimal blood pressure at the start of the study, developed hypertension within five years (Schutte et al., 2012; Hamer et al., 2015). A wide range of clinical and phenotypical analysis data are available for this cohort, including the golden standard 24 h ambulatory blood pressure monitoring (24 h ABPM) measurements and a measurement of hyperglycaemia (HbA1c, glycated haemoglobin), which are routinely used to identify hypertension and diabetes respectively in a clinical setup.

Using this unique cohort, we aim to determine if mtDNA variation, using the MutPred adjusted load as defining parameter, is different in those with hypertension or hyperglycaemia when compared to those without. As such, we present a new and updated method for attempting to associate mtDNA variation with a complex trait. This method would be applicable to all mtDNA association studies and be less affected by population stratification.

Table 1
Phenotypical data on SABPA cohort.

Parameter	Haplogroup L male	Haplogroup L female	Haplogroup MN male	Haplogroup MN female
Number of participants (n)	100	94	75	94
Age (year)	43.2 ± 8.17	45.7 ± 7.86	45.4 ± 10.5	44.5 ± 10.7
Body mass index	27.6 ± 5.77	32.9 ± 7.23	29.0 ± 5.33	26.2 ± 5.74
HbA1c (%)	6.23 ± 1.23	5.90 ± 1.14	5.66 ± 0.49	5.36 ± 0.29
24 h systolic ABPM (mmHg)	137.7 ± 16.0	128.6 ± 15.2	128.1 ± 10.6	120.7 ± 12.7
24 h diastolic ABPM (mmHg)	87.9 ± 10.7	78.8 ± 8.50	79.7 ± 7.72	73.8 ± 7.89

For each parameter, the group mean value is given ± standard deviation. Body mass index is a measure of body fat based on height and weight. HbA1c, glycated haemoglobin is measured to calculate the three month average plasma glucose concentration. 24 h ABPM, 24 h ambulatory blood pressure monitoring.

2. Results

After some samples fall-out due to technicalities such as sample unavailability and insufficient DNA extraction, 194 participants with macro-haplogroup L (haplogroups L0–L4) and 169 participants with macro-haplogroup MN (haplogroups M, N, R, B, H, I, J, K, T, U and W) were used in this study. In Table 1, some of the most important phenotypical measurements (age, body mass index, blood glucose levels and blood pressure) are summarised. Immediately, large differences in both systolic and diastolic 24 h ABPM measurements, as well as differences in blood glucose measurements, can be noted between different mtDNA background and gender groups (Fig. 1). For both blood pressure measurements (systolic and diastolic), haplogroup L males have significantly higher values than all other groups ($P < 0.0001$). Haplogroup L females also had significantly higher blood pressure than haplogroup MN females ($P = 0.001$ for both systolic and diastolic 24 h ABPM). Haplogroup MN males also had significantly higher blood pressure than their female counterparts ($P = 0.003$ for systolic 24 h ABPM; $P < 0.0001$ for diastolic 24 h ABPM). For blood glucose measurements (% of HbA1c) haplogroup L males had significantly higher values than both male and female groups from haplogroup MN ($P < 0.0001$), and only slightly higher values than haplogroup L females ($P = 0.03$); haplogroup L females had significantly higher blood glucose levels than their haplogroup MN counterparts ($P = 0.0004$) (Fig. 1). These differences have been previously reported in the current cohort (Lammertyn et al., 2011; Hamer et al., 2015), and correlates well with similar discrepancies between population groups reported in other studies (reviewed by Mensah, 2013). Consequently, data were divided into four groups based on gender and mtDNA background and analysed separately for subsequent analysis. These groups were as follows: haplogroup L males ($n = 100$); haplogroup L females ($n = 94$); haplogroup MN males ($n = 75$); haplogroup MN females ($n = 94$).

2.1. Functional network analysis

We performed network analyses on our data, using only variants with MutPred scores above 0.5, as well as rRNA and tRNA variants. Separate and simplified networks, shown in Fig. 2, were calculated for macro-haplogroups L and MN. Several occurrences of homoplasies were found in both macro-haplogroups (indicated in italics). This is not uncommon within the central regions of networks (Hernstadt et al., 2002). Activating the "frequency > 1" criterion, which then excludes all unique taxa, contributed greatly to simplifying the networks, but also resulted in the only participant from haplogroup L4 being excluded from Fig. 2A. The included nodes formed into clusters that correlated very well with the haplogroups that were assigned independently to each participant using the Haplogrep 2.0 program (Weissensteiner et al., 2016). It is

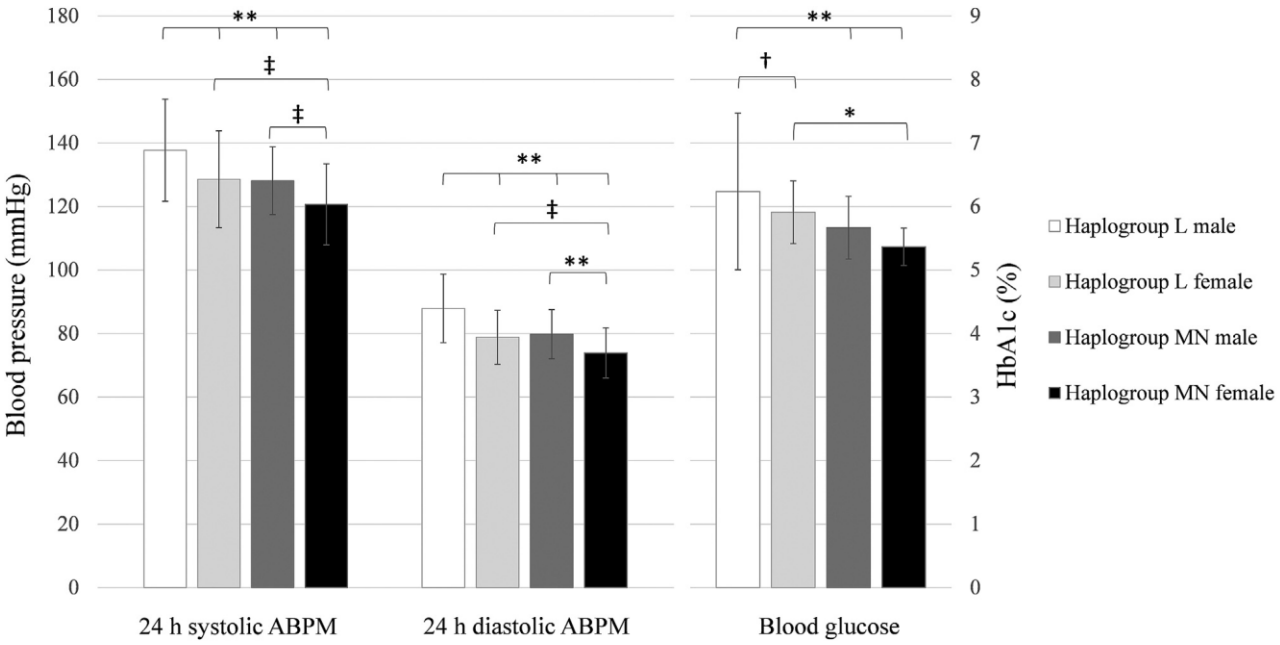


Fig. 1. Comparison of differences in systolic and diastolic 24 h ambulatory blood pressure monitoring measurements (ABPM), and glycated haemoglobin (HbA1c) as representation of blood glucose level, between gender and mtDNA background groups. ** $P < 0.0001$, * $P < 0.001$, ‡ $P < 0.01$, † $P < 0.05$.

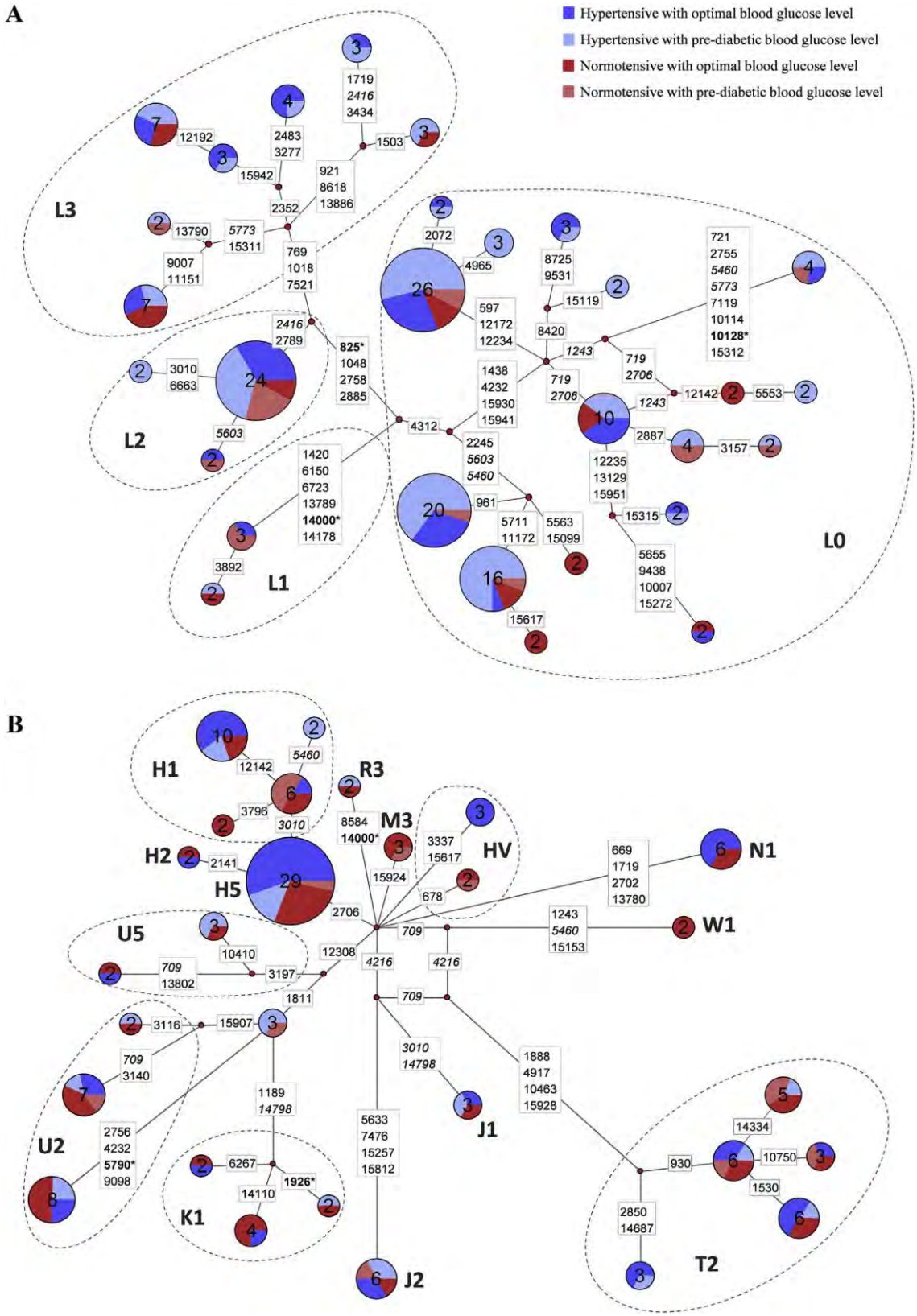


Fig. 2. Networks of haplogroup L (**A**) and haplogroup MN (**B**) participants. Ellipses indicate clusters of branches forming sub-haplogroups. Red nodes indicate potential median vectors. Coloured pie-chart nodes indicate the proportion of individuals per phenotype classification, as well as the number of participants per node. mtDNA variants responsible for links are listed in boxes; homoplasic variants are shown in italics, while transversions are shown in bold and marked with an *.

interesting to note that the phenotype classifications do not group within specific parts of the network and seem well dispersed throughout the networks (Fig. 2).

2.2. mtDNA variants

We found very few heteroplasmies over 10% and therefor did not consider the possible contribution of these sites to the phenotypes investigated here. We also found no clinically proven mutations at appreciable heteroplasmies among any of our participants. In our current dataset, we found 17 changes not present in GenBank sequences. Of those, twelve changes were transitions and five changes were transversions. Four of these changes have however been previously reported, according to the MITOMAP database. The properties of these changes are listed in Table 2.

To capture any differences in the population frequency of rare population alleles, we compared the number of rare (<0.1%) and common (>1%) non-synonymous population alleles unique to each group as defined by haplogroup background (L or MN), gender and phenotype (in this analysis, a two-group classification for blood glucose levels were used instead of three groups, i.e., optimal or pre-diabetic glucose levels, using a cut-off of HbA1c = 5.7%). Although rare population allele counts were higher in both male and female haplogroup L hypertensive groups, as well as in the haplogroup L pre-diabetic group, these differences did not prove to be statistically significant. Possibly, the amount of rare population alleles identified in these groups is influenced by the under-representation of African sequences in the GenBank database, or the disproportionate number of haplogroup L participants that are classified as hypertensive (about 80%). As can be seen in Table 3, no statistically significant differences were found between any of the other groups when looking at either hypertension or hyperglycaemia.

2.3. Mutational load calculations and adjustments using MutPred pathogenicity scores

Next, MutPred adjusted loads, which serve as an estimate of the likely impact of a person's non-synonymous mtDNA substitutions on the functioning of their mtDNA encoded proteins, were calculated as described in the Introduction. Since i) the possible pathogenicity of variants with a MutPred score greater than 0.5 can be considered an "actionable hypothesis" (Li et al., 2009), and ii) these higher scoring variants are thought to be rarer and less likely to be common population variants (Pereira et al., 2011), we applied specific MutPred score criteria while carrying out statistical analyses. Table S1 shows the sum of MutPred scores (MutPred mutational load), the number of MutPred scored variants, and the calculated MutPred adjusted load for each participant, using these different criteria. Table 4 summarises the data by giving the means and standard deviations for the above mentioned genetic parameters, when using either all MutPred scored variants, or only those with MutPred scores above the 0.5 threshold. The impact of the reference sequence location is greatly reduced when only variants with MutPred scores greater than 0.5 are considered. As shown in Fig. 3, doing this resulted in a 37% change in calculated MutPred adjusted loads in haplogroup L groups, but only a 13% change in haplogroup MN groups, underlining the utility of applying this type of correction in studies using diverse populations. As previously mentioned, variants with low MutPred scores are likely to be close to neutral; we do not expect these variants to impact on phenotype. This is supported by the lack of purification selection against these variants (Cavadas et al., 2015). Our own data also reflected this, as haplogroup L participants had an average of 10.2 variants with MutPred scores below 0.5, while haplogroup MN participants only had an average of 4.0 low scoring variants. This ratio is greatly reduced when moving to the other side of the MutPred scale: haplogroup L participants had an average of 1.6 variants with MutPred scores above 0.5, while haplogroup MN participants had an average of 1.1 variants with scores above this threshold. Excluding these variants (MutPred scores less than 0.5) from further statistical analyses that followed then also allowed us to minimise their impact on our genetic parameter (MutPred adjusted load). The large standard deviations present in Fig. 3, when only variants with MutPred scores above 0.5 are included, can be attributed to the fact that 37% of participants (45 from haplogroup L and 89 from haplogroup MN) do not have any variants above this threshold.

2.4. Correlations between MutPred adjusted loads and measurements of blood pressure and blood glucose levels

To determine if there is any relationship between MutPred adjusted loads and measurements of blood pressure (Table 5) or

Table 2
A list of novel variants found in the current cohort.

rCRS position	rCRS NT	Variant NT	Mutation type	Gene	Mutation effect	MITOMAP report	Conservation (%)	MutPred pathogenicity score
8423	C	T	Transition	*MT-ATP8*	syn: L→L	–	88.9	–
7351	T	C	Transition	*MT-COI*	non-syn: L→P	–	48.9	0.432
7291	C	T	Transition	*MT-COI*	non-syn: T→M	–	86.7	0.353
14945	G	A	Transition	*MT-CYB*	non-syn: A→T	–	40.0	0.523
15577	C	T	Transition	*MT-CYB*	syn: A→A	–	97.8	–
15124	A	G	Transition	*MT-CYB*	syn: T→T	Single published	97.8	–
14718	A	G	Transition	*MT-TE*	tRNA	–	2.2	–
4842	A	G	Transition	*MT-ND2*	non-syn: T→A	Single unpublished	35.6	0.386
5106	A	G	Transition	*MT-ND2*	non-syn: T→A	–	20.0	0.379
4479	C	T	Transition	*MT-ND2*	syn: L→L	–	46.7	–
13552	G	A	Transition	*MT-ND5*	non-syn: A→T	–	84.4	0.513
13538	T	C	Transition	*MT-ND5*	non-syn: M→T	–	20.0	0.341
2603	C	A	Transversion	*MT-RNR2*	rRNA	–	100.0	–
6320	T	G	Transversion	*MT-COI*	syn: P→P	–	22.2	–
5232	C	A	Transversion	*MT-ND2*	non-syn: P→T	Single unpublished	100.0	0.780
4906	C	A	Transversion	*MT-ND2*	non-syn: S→Y	–	53.3	0.535
12969	C	G	Transversion	*MT-ND5*	syn: T→T	Single unpublished	8.9	–

A list of variants found in this cohort, but not present on GenBank is given. The nucleotide (NT) change for each position is given, followed by descriptive information about the variant. Four variants have been reported according to MITOMAP records, and three are from unpublished records. rCRS, revised Cambridge Reference Sequence; syn, synonymous change; non-syn, non-synonymous change.

Table 3
Number of non-synonymous variants unique to each group.

GenBank frequency (%)	HT group	NT group	*P*-value	PD-BG group	O-BG group	*P*-value
Haplogroup L male						
<0.1	18	3	1	13	4	1
>1	5	1		4	1	
Haplogroup L female						
<0.1	14	2	1	6	10	0.19
>1	9	2		1	9	
Haplogroup MN male						
<0.1	1	4	0.28	2	2	0.52
>1	6	4		2	9	
Haplogroup MN female						
<0.1	8	3	0.38	2	8	1
>1	4	4		2	7	

The number of non-synonymous variants were calculated according to set criteria: very rare variants were those that are present in less than 0.1% of GenBank sequences; common variants are those that are present in more than 1% of GenBank sequences. Only non-synonymous variants that are unique for a specific group as defined by haplogroup, gender and disease classification were counted. *P*-values for a two-tailed Fisher's exact test are given next to counts. HT, hypertensive; NT, normotensive; PD-BG, pre-diabetic blood glucose levels; O-BG, optimal blood glucose levels.

Table 4
Summary of mtDNA variant parameters.

Group	MutPred mut. load per par.		Number of var. per par.		MutPred adj. load per par.	
	All var. included	Var. with MutPred scores > 0.5	All var. included	Var. with MutPred scores > 0.5	All var. included	Var. with MutPred scores > 0.5
Haplogroup L male	3.94 ± 1.50	0.85 ± 0.80	11.5 ± 3.6	1.4 ± 1.3	0.33 ± 0.04	0.44 ± 0.26
Haplogroup L female	4.18 ± 1.58	1.02 ± 0.99	12.1 ± 3.6	1.7 ± 1.7	0.34 ± 0.04	0.48 ± 0.23
Haplogroup MN male	2.04 ± 1.01	0.70 ± 0.72	5.3 ± 2.4	1.1 ± 1.2	0.39 ± 0.05	0.34 ± 0.31
Haplogroup MN female	1.94 ± 0.98	0.64 ± 0.66	4.9 ± 2.5	1 ± 1.1	0.40 ± 0.04	0.35 ± 0.31

The means ± standard deviation are given for specified genetic parameters per group. MutPred mutational load is the sum of MutPred scores for either all variants per participant, or only for variants with a MutPred score above 0.5. Number of variants refers to those variants with a MutPred score. The MutPred adjusted load is the MutPred mutational load per participant divided by the number of variants per participant. MutPred, a pathogenicity score assigned to a variant using the MutPred program (mutpred.mutdb.org/about.html); mut., mutational; var., variants; par., participant; adj., adjusted.

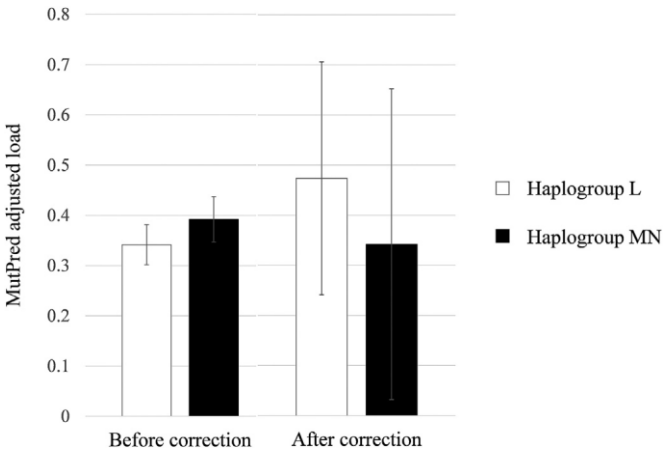


Fig. 3. Representation of the effect on MutPred adjusted load when a correction is applied to address population stratification.

blood glucose (Table 6), Pearson's correlation analyses were carried out for each of the four gender/mtDNA background groups. From Table 5, it can be seen that several MutPred adjusted loads were significantly correlated with blood pressure measurements. However, none of these significant associations remained after correcting for multiple comparisons. From Table 6, it can be seen that none of the MutPred adjusted loads, for either the entire mtDNA molecule, or those of individual electron transport chain (ETC) complexes, correlated significantly with blood glucose measurements.

2.5. Comparison of MutPred scores between different classifications of blood pressure and blood glucose levels

Next, we classified participants into different groups within the two CVD phenotypes that were investigated (hypertension and hyperglycaemia) as described in Materials and methods (Table S1). To investigate the role of mtDNA variants in hypertension, we compared the mean values for genetic parameters of normotensive and hypertensive participants using independent *t*-tests. From Table 7, it is apparent that the total MutPred adjusted load differed significantly between normotensives and hypertensives within haplogroup L females, as well as haplogroup MN males. Similar to the above associations, after correction for multiple comparisons, significance was lost. We then compared the mean values for genetic parameters of participants with optimal blood glucose levels, those classified as pre-diabetic and those classified as diabetic within the different gender/mtDNA background groups, using one-way ANOVAS. None of these comparisons differed significantly (results not shown).

2.6. Comparison of high MutPred scoring variant frequency between different classifications of blood pressure and blood glucose levels

To see if having one or more variants with very high MutPred pathogenicity scores can increase the risk of disease, we determined the frequency of these high scoring variants within the different classifications in our CVD phenotypes. The number of participants with one or more variants with MutPred scores above 0.6, 0.7 or 0.8 within each gender/mtDNA background group is

Table 5
Pearson's correlation coefficients for genetic parameters compared to mean values of 24 h ABPM measurements.

Locus	24 h ABPM	Total cohort ($n = 363$)		Haplogroup L male ($n = 100$)		Haplogroup L female ($n = 94$)		Haplogroup MN male ($n = 75$)		Haplogroup MN female ($n = 94$)	
		MutPred	Pearson r	MutPred	Pearson r	MutPred	Pearson r	MutPred	Pearson r	MutPred	Pearson r
Total mtDNA	Systolic	0.41 ± 0.28	.096	0.43 ± 0.25	.234*	0.48 ± 0.23	–.063	0.33 ± 0.31	–.110	0.34 ± 0.30	.156
	Diastolic		.075		.233*		–.067		–.241*		.122
CI	Systolic	0.37 ± 0.28	.051	0.41 ± 0.26	.110	0.45 ± 0.24	–.062	0.30 ± 0.30	–.073	0.28 ± 0.30	.011
	Diastolic		.040		.128		–.050		–.198		–.052
CIII	Systolic	0.09 ± 0.22	**–.105***	0.05 ± 0.17	–.018	0.07 ± 0.19	–.177	0.10 ± 0.23	–.015	0.12 ± 0.25	.037
	Diastolic		**–.129***		–.007		–.113		–.217		.001
CIV	Systolic	0.05 ± 0.17	**.117***	0.07 ± 0.19	**.217***	0.06 ± 0.19	–.014	0.01 ± 0.09	–.082	0.03 ± 0.13	**.271****
	Diastolic		**.116***		.154		–.009		**–.253***		**.358****
CV	Systolic	0.07 ± 0.20	.067	0.05 ± 0.18	**.271****	0.04 ± 0.16	**.234***	0.05 ± 0.19	–.205	0.07 ± 0.21	.161
	Diastolic		.009		.182		.093		–.164		.099

Pearson's correlations were carried out to investigate the relationship between MutPred adjusted loads for mtDNA loci as indicated, and blood pressure measurements. Means ± standard deviations of MutPred adjusted loads are given with Pearson's correlation coefficients (r) for systolic and diastolic 24 h ABPM respectively. Correlations of significance (before correction for multiple testing) are indicated in bold and by * (significant at the 0.05 level) and ** (significant at the 0.01 level). CI, NADH:ubiquinone oxidoreductase; CIII, cytochrome c oxidoreductase; CIV, cytochrome c oxidase; CV, ATP synthase.

Table 6
Pearson's correlation coefficients for genetic parameters compared to blood glucose measurements.

Locus	Total cohort ($n = 363$)		Haplogroup L male ($n = 100$)		Haplogroup L female ($n = 94$)		Haplogroup MN male ($n = 75$)		Haplogroup MN female ($n = 94$)	
	MutPred	Pearson r	MutPred	Pearson r	MutPred	Pearson r	MutPred	Pearson r	MutPred	Pearson r
Total mtDNA	0.41 ± 0.28	.039	0.43 ± 0.25	.062	0.48 ± 0.23	–.003	0.33 ± 0.31	–.076	0.34 ± 0.30	–.147
CI	0.37 ± 0.28	.048	0.41 ± 0.26	.013	0.45 ± 0.24	.019	0.30 ± 0.30	–.045	0.28 ± 0.30	–.164
CIII	0.09 ± 0.22	–.064	0.05 ± 0.17	–.002	0.07 ± 0.19	.035	0.10 ± 0.23	–.099	0.12 ± 0.25	.006
CIV	0.05 ± 0.17	.035	0.07 ± 0.19	.115	0.06 ± 0.19	–.040	0.01 ± 0.09	–.078	0.03 ± 0.13	–.069
CV	0.07 ± 0.20	–.005	0.05 ± 0.18	–.040	0.04 ± 0.16	.181	0.05 ± 0.19	–.041	0.07 ± 0.21	.088

Pearson's correlations were carried out to investigate the relationship between MutPred adjusted loads for mtDNA loci as indicated, and blood glucose (% of HbA1c) measurements. Means ± standard deviations of MutPred adjusted loads are given with Pearson's correlation coefficients (r). CI, NADH:ubiquinone oxidoreductase; CIII, cytochrome c oxidoreductase; CIV, cytochrome c oxidase; CV, ATP synthase.

Table 7
Mean MutPred adjusted loads of normotensive participants compared to those of hypertensive participants.

Locus	Blood pressure status	Total cohort (NT = 115, HT = 248)		Haplogroup L male (NT = 22, HT = 78)		Haplogroup L female (NT = 18, HT = 76)		Haplogroup MN male (NT = 33, HT = 42)		Haplogroup MN female (NT = 42, HT = 52)	
		MutPred	t (df)	MutPred	t (df)	MutPred	t (df)	MutPred	t (df)	MutPred	t (df)
Total mtDNA	Normal	0.41 ± 0.29	–0.06 (384)	0.32 ± 0.30	1.97 (28)	0.59 ± 0.04	–4.67 (89)**	0.41 ± 0.30	–2.01 (73)*	0.33 ± 0.31	0.24 (92)
	Hyper	0.40 ± 0.27		0.47 ± 0.23		0.45 ± 0.24		0.27 ± 0.30		0.35 ± 0.30	
CI	Normal	0.36 ± 0.29	0.21 (384)	0.32 ± 0.30	1.62 (29)	0.56 ± 0.14	**–2.84 (46)****	0.36 ± 0.30	–1.5 (–73)	0.28 ± 0.30	0.03 (92)
	Hyper	0.37 ± 0.28		0.43 ± 0.25		0.43 ± 0.26		0.26 ± 0.30		0.28 ± 0.30	
CIII	Normal	0.12 ± 0.25	–1.59 (209)	0.05 ± 0.18	0.10 (98)	0.13 ± 0.25	–1.15 (–20)	0.14 ± 0.27	–1.2 (–57)	0.11 ± 0.25	0.45 (92)
	Hyper	0.08 ± 0.20		0.05 ± 0.17		0.05 ± 0.17		0.07 ± 0.20		0.13 ± 0.26	
CIV	Normal	0.05 ± 0.18	–0.27 (384)	0.06 ± 0.19	0.30 (98)	0.15 ± 0.29	–1.48 (–19)	0.03 ± 0.14	–1.4 (–32)	0.01 ± 0.09	1.15 (85)
	Hyper	0.05 ± 0.16		0.07 ± 0.19		0.04 ± 0.15		0 ± 0		0.04 ± 0.16	
CV	Normal	0.07 ± 0.21	–0.30 (384)	0.02 ± 0.13	0.83 (98)	0 ± 0	**2.75 (75)****	0.09 ± 0.24	–1 (–45)	0.07 ± 0.21	–0.03 (92)
	Hyper	0.06 ± 0.19		0.06 ± 0.19		0.05 ± 0.17		0.02 ± 0.13		0.07 ± 0.21	

Independent t-tests were carried out to compare means of MutPred adjusted loads for mtDNA loci as indicated, between normotensive (NT) and hypertensive (HT) participants. Means ± standard deviations of MutPred adjusted loads are given with the test statistic (t) and degrees of freedom (df). Differences of significance (before correction for multiple testing) are indicated in bold and by * (significant at the 0.05 level) and ** (significant at the 0.01 level). CI, NADH:ubiquinone oxidoreductase; CIII, cytochrome c oxidoreductase; CIV, cytochrome c oxidase; CV, ATP synthase.

listed in Table 8 (hypertension classifications) and Table 9 (hyperglycaemia classifications). Also, in Fig. 4, the average number of variants per person belonging to a specific macro-haplogroup is represented. It can be seen that haplogroup L participants are no more likely than haplogroup MN participants to have one or more variants with a MutPred score above 0.6; haplogroup MN participants are slightly more likely to have at least one variant with a MutPred score above 0.7. For variants with MutPred scores above 0.8, the average number of instances is higher in haplogroup L participants compared to haplogroup MN participants (0.03 *vs.* 0.005), but these refer to a very low number of counts. This data correlate well with previous reports by Pereira et al. (2011) and Cavadas et al. (2015), who demonstrated that high MutPred scoring variants are more numerous in more recent branches (here haplogroup MN) of the phylogenetic tree, than branches deeper within the phylogenetic tree (here haplogroup L), due to purification selection. However, we found no significant differences in the number of participants carrying high MutPred scoring variants when comparing normotensives with hypertensives in any of the groups. We also found no significant differences between the three hyperglycaemia classifications within any of the groups.

Table 8
Number of participants within each gender, mtDNA background and blood pressure phenotype group having one or more variants with a MutPred score above 0.6, 0.7 or 0.8.

Number of variants	Number of haplogroup L male		Number of haplogroup L female		Number of haplogroup MN male		Number of haplogroup MN female	
	NT ($n = 22$)	HT ($n = 78$)	NT ($n = 18$)	HT ($n = 76$)	NT ($n = 33$)	HT ($n = 42$)	NT ($n = 42$)	HT ($n = 52$)
MutPred score > 0.6								
1	8	32	11	36	4	5	8	9
2	4	10	2	6	15	11	10	18
3	0	5	2	4	2	1	2	0
4	0	0	1	1	0	0	0	0
MutPred score > 0.7								
1	2	6	2	3	7	3	4	5
2	0	0	1	0	0	0	0	0
MutPred score > 0.8								
1	1	2	0	1	1	0	0	0
2	0	0	1	0	0	0	0	0

The number of participants in a specified group (e.g., hypertensive haplogroup L males) that have one or more variants (as indicated in the left first column) with a MutPred pathogenicity score above 0.6, 0.7 or 0.8 is given. Normotensive (NT) and hypertensive (HT) classifications are based on systolic and diastolic 24 h ABPM.

Table 9
Number of participants within each gender, mtDNA background and blood glucose phenotype group having one or more variants with a MutPred score above 0.6, 0.7 or 0.8.

Number of variants	Number of haplogroup L male			Number of haplogroup L female			Number of haplogroup MN male			Number of haplogroup MN female		
	O-BG ($n = 30$)	PD-BG ($n = 51$)	D-BG ($n = 19$)	O-BG ($n = 45$)	PD-BG ($n = 41$)	D-BG ($n = 8$)	O-BG ($n = 45$)	PD-BG ($n = 27$)	D-BG ($n = 3$)	O-BG ($n = 77$)	PD-BG ($n = 16$)	D-BG ($n = 1$)
MutPred score > 0.6												
1	9	23	8	18	21	7	6	3	0	15	1	0
2	3	9	2	5	3	0	14	11	1	24	4	0
3	2	2	1	3	3	0	2	1	0	2	0	0
4	0	0	0	1	1	0	0	0	0	0	0	0
MutPred score > 0.7												
1	2	4	2	3	2	0	5	5	0	8	1	0
2	0	0	0	0	1	0	0	0	0	0	0	0
MutPred score > 0.8												
1	1	2	0	1	0	0	0	1	0	0	0	0
2	0	0	0	0	1	0	0	0	0	0	0	0

The number of participants in a specified group (e.g., African males with optimal blood glucose levels) that have one or more variants with a MutPred pathogenicity score above 0.6, 0.7 or 0.8 is given. HbA1c measurements were used to classify participants as having either optimal blood glucose levels (O-BG, optimal blood glucose levels, HbA1c < 5.7%), being prediabetic (PD-BG, prediabetic blood glucose levels, HbA1c from 5.7% to 6.4%) or diabetic (D-BG, diabetic blood glucose levels, HbA1c > 6.4%).

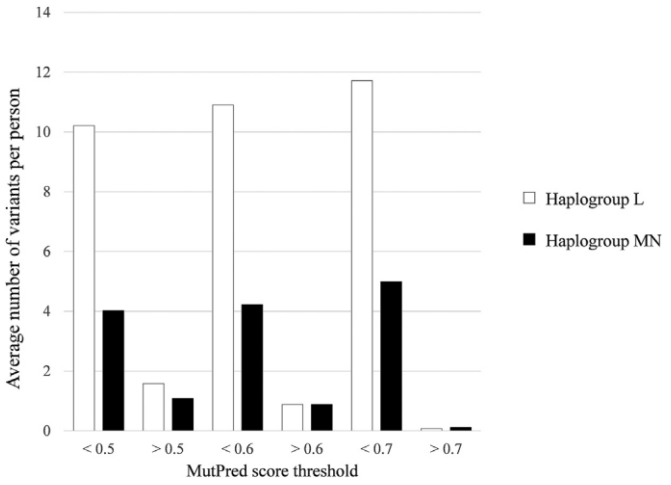


Fig. 4. The average number of variants with MutPred scores above different thresholds, per person within a specific haplogroup (L or MN).

3. Discussion

From the landmark Framingham Heart Studies (Mahmood et al., 2014), several risk factors have been identified as significant predictors of hypertension. In the SABPA cohort also, these factors, which included gender and population group, significantly predicted blood pressure (Hamer et al., 2015; Malan et al., 2015). Disparities in CVD onset and development between different population groups are often linked to gross socio-economic inequality (Kaufman et al., 2015). However, it is important to note that even though participants were matched on several factors, including socio-economic status, this disparity is still present in the SABPA cohort. In this study also, we showed that both gender and background contributed to significant differences in measurements of blood pressure and blood glucose. Consequently, the cohort was divided into four gender/mtDNA background groups and analysed separately as has been done in other studies using the SABPA participants.

Assuming that rare non-synonymous population alleles (present in <0.1% of GenBank sequences) are more likely to be deleterious than common population alleles (present in >1% of GenBank sequences), one would expect rare population alleles to be present more frequently in the affected members of the cohort if mildly

deleterious mtDNA variants were affecting susceptibility to disease. That was however not the case in our cohort, as no significant differences in the numbers of these population alleles were identified when the groups were compared. This is a less sophisticated analysis however, as the likelihood of pathogenicity of each allele is assumed, but not qualified by the use of more robust criteria. Using the MutPred program to qualify the pathogenicity of variants enabled us to carry out more sophisticated analyses. The usefulness of the MutPred scoring system to predict pathogenicity has been thoroughly demonstrated by Pereira et al. (2011) and Cavadas et al. (2015).

Genetic parameters such as the MutPred mutational load per participant, the number of MutPred scored variants per participant, and the MutPred adjusted load per participant, were calculated using specified variant inclusion criteria. Because of the position of the revised Cambridge Reference Sequence (rCRS) in the phylogeny, African lineages are expected to have more variants than European lineages which lie closer on the phylogeny to the reference. Importantly, many of these variants have MutPred scores below 0.5, are common population polymorphisms, and thus have a low likelihood of being mildly deleterious. The exclusion of variants with MutPred scores below 0.5 reduced the differences in the mean number of variants between haplogroup L and MN groups. Interestingly, it also increased the mean MutPred adjusted loads in haplogroup L groups, while it decreased this mean in haplogroup MN groups (Table 4). However, some common haplogroup defining variants do have MutPred scores above 0.5, and are still included (Table S2). It is for this reason that a second step of correction is needed, i.e., adjusting mutational load scores according to sequence position within the phylogeny. By excluding these low-impact variants from analyses, and then calculating MutPred adjusted loads, any bias that is potentially introduced by phylogenetic distance, when utilizing mutational load methods, is largely corrected.

We investigated the role of mtDNA variants in the presentation (or severity) of two major CVD phenotypes (hypertension and hyperglycaemia). MutPred adjusted loads for variants with MutPred pathogenicity scores above 0.5 were used in correlation tests. Although some significant differences were found when comparing MutPred adjusted loads between hypertensives and normotensives, these significances were not robust enough to withstand corrections for multiple testing. No significant differences between any groups were found within the hyperglycaemia classifications.

Although haplogroup L participants were more likely to have at least one high MutPred scoring variant above 0.5 than haplogroup MN participants, they were no more likely to have very high MutPred scoring variants (above 0.7). We found no significant relationship between the prevalence of potentially deleterious variants and any phenotypical classification.

Thus, from our data, we could not identify a relationship between MutPred adjusted loads or high MutPred scoring variants and the presentation of our investigated phenotypes. We however did not investigate the role of these genetic parameters in the outcome of disease. It is also important to note that this study was not sufficiently powered to detect differences in individual ETC complexes, which would provide a much more refined outcome. We did however present a new and updated method for attempting to associate mtDNA variation with a complex trait.

The current study is one of very few mtDNA genetic studies conducted in a cohort that includes African participants with extensively phenotyped CVD. The global burden of CVDs and the rising burden thereof in Africa make this collection of phenotypes especially important to investigate. More rare non-synonymous population alleles (present in <0.1% of GenBank sequences) were identified in haplogroup L participants than haplogroup MN participants, possibly influenced by the low presence of African sequences in the GenBank database. This underscores again the need to sequence more mtDNA from African populations, and especially from under-represented Southern African population groups.

Despite our data not indicating a statistically significant link between mtDNA variation and the two phenotypes investigated here, we are of the opinion that the method proposed here delivered well-powered results for the specific questions we posed, especially considering the modest size of the cohort. With the validity of the MutPred system being well established, we suggest the Mutational adjusted load method proposed here could be useful in similar investigations on other phenotypes and diseases. These investigations should indeed be plentiful in the near future, with many associations between mtDNA variation and common complex disease being proposed.

4. Materials and methods

4.1. Cohort recruitment and sample collection

We included participants of the SABPA cohort in this study. Details on participant recruitment, and sample and data collection have been published elsewhere (Hamer et al., 2015; Malan et al., 2015). Inclusion criteria for the SABPA study were urban Black and Caucasian male and female teachers from South Africa ($n = 409$), with similar socio-economic status, aged 20–65 years, from the North-West Province (Malan et al., 2015). Participants were enrolled in the project in 2008–2009, having their first follow up in 2011–2012. During recruitment and before consent, study participants were informed about the objectives and procedures of the study, including investigations of genes involved in CVD. The study complied with criteria on human research set by the Helsinki declaration (2004) and was approved by the Ethics Committee of the North-West University (NWU 00036-07-S6).

4.2. Sequencing and data management

The complete mtDNA genome of each participant was sequenced at an average base coverage of about 1800× from two overlapping long range PCR products as previously described (van der Walt et al., 2012), using the Ion PGM and 200-bp templating and sequencing chemistry. This method was validated using two additional sequence chemistries (Illumina and pyrosequencing, QIAGEN, USA). Base calling and alignment were automatically performed by the Ion Torrent Suite 4.4 software. Variants were identified by the VariantCaller plug-in (version 5.0) using the rCRS (Andrews et al., 1999). A list of variants for each participant can be found in the supplementary file "JGG Supplementary variants SABPA". Only consensus variants were used in this study. We did not investigate the effect of heteroplasmy here. Haplogroups were assigned using the online Haplogrep 2.0 tool (haplogrep.uibk.ac.at). MutPred scores were assigned to all non-synonymous variants (mutpred.mutdb.org/about.html). MutPred adjusted loads were calculated by obtaining the average of N, where N is the set of MutPred score values for variants within a specific target region, i.e., a single gene, a set of genes within an enzyme complex or the entire mtDNA genome. A functional maximum parsimony (MP) network analysis was performed with the NETWORK (version 5) and NETWORK Publisher (version 2.0.0.1) software packages (http://www.fluxus-engineering.com/sharenet.htm), by using only mtDNA variants with MutPred pathogenicity scores above 0.5, rRNA variants and tRNA variants. The processing strategy to produce the networks, were aimed at reducing complexity (introduced by the large amount of

sequences). Transversions, being chemically less likely to occur, were weighted three times more than transitions. Reduced median (RM) (Bandelt et al., 1995) processing split loci on the basis of genetic distance (reduction threshold r was 3 and 2 for macrohaplogroup L and MN networks respectively). This step was followed by median joining (MJ) (Bandelt et al., 1999) to calculate all possible shortest trees (ε was set to 30 and 0 for macrohaplogroup L and MN networks, respectively) and produce a network. The "frequency > 1" criterion was activated to ignore unique sequences/taxa and reduce complexity. Finally, an MP clean-up step (Polzin and Daneschmand, 2003) was performed to remove unnecessary median vectors and links.

All figures were produced using Excel (Microsoft Office 365) or NETWORK Publisher (version 2.0.0.1) and were finalised in PowerPoint (Microsoft Office 365) and Adobe Photoshop Elements (Version 10.0).

4.3. Phenotyping

Normotensive and hypertensive classifications are based on 24 h ABPM measurements. A hypertensive classification was assigned to males with 24 h systolic ABPM $\geq$ 130 mmHg and/or 24 h diastolic ABPM $\geq$ 80 mmHg, and females with 24 h systolic ABPM $\geq$ 125 mmHg and/or 24 h diastolic ABPM $\geq$ 75 mmHg (Hermida et al., 2013; Mancia et al., 2013). HbA1c expressed as a percentage, which is measured to calculate the three month average plasma glucose concentration, was used to classify participants as having either optimal blood glucose levels (HbA1c below 5.7%), falling within pre-diabetic ranges (HbA1c from 5.7% to 6.4%) or diabetic ranges (HbA1c above 6.4%) (American Diabetes Association, 2010). For a two-group classification (optimal or pre-diabetic glucose levels), a cut-off value of HbA1c = 5.7% was used.

4.4. Statistical analysis

All statistical analyses were carried out using IBM SPSS Statistics (version 23) or GraphPad Prism software (version 6.05). A one-way analysis of variance (ANOVA) followed by a post-hoc Tuckey test, was carried out to assess the relationships between gender and mtDNA background and two CVD clinical measurements (blood pressure and blood glucose levels).

The population frequency of rare population alleles, defined as having a frequency (GenBank frequency) below 0.1% on the global databases of more than 30,000 sequences (MITOMAP: A human mitochondrial genome database. http://www.mitomap.org), was compared between those with optimal and those with high blood pressure or high blood glucose levels using a Fisher's exact test.

To assess the number of participants in each phenotype classification who possess one or more variants with MutPred pathogenicity scores above 0.6, 0.7 or 0.8, we applied Fisher's exact tests; a Bonferroni correction for multiple testing was applied to both sets of comparisons.

Pearson's correlation analyses were used to assess the correlation between MutPred adjusted loads, blood pressure and blood glucose levels in each of the four gender/mtDNA background groups, using a Bonferroni correction for multiple comparisons.

Mean MutPred adjusted loads of normotensive and hypertensive participants in the four groups were compared using independent t-tests, with a Bonferroni correction for multiple comparison. For participants with optimal blood glucose levels, those within pre-diabetic ranges, and those within diabetic ranges, mean MutPred adjusted loads were compared using one-way ANOVA, with a Tukey post-hoc test.

Acknowledgments

We acknowledge the Faculty of Natural Sciences of the North-West University for contributing to funding and Thermo Fisher South Africa for providing additional technical resources to this study. The present study was partially funded by the South African National Research Foundation; Medical Research Council; ROCHE Diagnostics; North-West University, South Africa; as well as the Metabolic Syndrome Institute, France. We would also like to acknowledge funding support from the Royal Society and the National Research Foundation of South Africa, for the academic meeting at which the project was mapped out.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jgg.2016.12.003.

References

Achilli, A., Olivieri, A., Pala, M., Hooshiar Kashani, B.C., Perego, U.A., Gandini, F., Santoro, A., Battaglia, V., Grugni, V., Lancioni, H., Sirolla, C., Bonfigli, A.R., Cormio, A., Boemi, M., Testa, I., Semino, O., Ceriello, A., Spazzafumo, L., Gadaleta, M.N., Marra, M., Testa, R., Franceschi, C., Torroni, A., 2011. Mitochondrial DNA backgrounds might modulate diabetes complications rather than T2DM as a whole. PLoS One 6, e21029.

Ameh, J., Godwin, I., Obi, I., Puepet, F., Aminu, B., Suleiman, T., 2011. The search for mitochondrial tRNALeu(UUR) A3243G mutation among type 2 diabetes mellitus patients in the Nigerian population. Afr. J. Biotechnol. 10, 13383–13389.

American Diabetes Association, 2010. Standards of medical care in diabetes. Diabetes Care 33 (Suppl. 1), S11–S61.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N., 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23, 147.

Bandelt, H.-J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48.

Bandelt, H.-J., Forster, P., Sykes, B.C., Richards, M.B., 1995. Mitochondrial portraits of human populations. Genetics 141, 743–753.

Cardena, M.M., Ribeiro-dos-Santos, A., Santos, S., Mansur, A.J., Pereira, A.C., Fridman, C., 2014. Amerindian genetic ancestry is associated with higher survival rates compared to African and European ancestry in Brazilian patients with heart failure. Int. J. Cardiol. 176, 527–528.

Cardena, M.M., Ribeiro-Dos-Santos, A.K., Santos, S.E., Mansur, A.J., Bernardez-Pereira, S., Santos, P.C., Pereira, A.C., Fridman, C., 2016. Mitochondrial and genomic ancestry are associated with etiology of heart failure in Brazilian patients. J. Hum. Hypertens. 30, 120–123.

Cavadas, B., Soares, P., Camacho, R., Brandao, A., Costa, M.D., Fernandes, V., Pereira, J.B., Rito, T., Samuals, D., Periera, L., 2015. Fine time scaling of purifying selection on human nonsynonymous mtDNA mutations based on the worldwide population tree and mother-child pairs. Hum. Mutat. 36, 1100–1111.

Chinnery, P.F., Elliott, H.R., Syed, A., Rothwell, P.M., Oxford vascular study, 2010. Mitochondrial DNA haplogroups and risk of transient ischaemic attack and ischaemic stroke: a genetic association study. Lancet Neurol. 9, 498–503.

Elson, J.L., Andrews, R.M., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N., 2001. Analysis of European mtDNAs for recombination. Am. J. Hum. Genet. 68, 145–153.

Elson, J.L., Herrnstadt, C., Preston, G., Thal, L., Morris, C.M., Edwardson, J.A., Beal, M.F., Turnbull, D.M., Howell, N., 2006. Does the mitochondrial genome play a role in the etiology of Alzheimer's disease? Hum. Genet. 119, 241–254.

Elson, J.L., Turnbull, D.M., Howell, N., 2004. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. Am. J. Hum. Genet. 74, 229–238.

Gorman, G.S., Schaefer, A.M., Ng, Y., Gomez, N., Blakely, E.L., Alston, C.L., Feeney, C., Horvath, R., Yu-Wai-Man, P., Chinnery, P.F., Taylor, R.W., Turnbull, D.M., McFarland, R., 2015. Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. Ann. Neurol. 77, 753–759.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., Ritchie, G.R., Xue, Y., Asimit, J., Nsubuga, R.N., Young, E.H., Pomilla, C., Kivinen, K., Rockett, K., Kamali, A., 2015. The African genome variation project shapes medical genetics in Africa. Nature 517, 327–332.

Hamer, M., von Känel, R., Reimann, M., Malan, N.T., Schutte, A.E., Huisman, H.W., Malan, L., 2015. Progression of cardiovascular risk factors in black Africans: 3 year follow up of the SABPA cohort study. Atherosclerosis 238, 52–54.

Hermida, R.C., Ayala, D.E., Mojon, A., Fontao, M.J., Chayan, L.F., 2013. Differences between men and woman in ambulatory blood pressure thresholds for diagnosis of hypertension based on cardiovascular outcomes. Chronobiol. Int. 30, 221–232.

Hernstadt, C., Elson, J.L., Fahy, E., Preston, G., Turnbull, D.M., Anderson, C.,

Ghosh, S.S., Olefsky, J.M., Beal, M.F., Davis, R.E., Howell, N., 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African Asian, and European haplogroups. Am. J. Hum. Genet. 70, 1152–1171.
Hernstadt, C., Howell, N., 2004. An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders. Mitochondrion 4, 791–798.
Hollingsworth, K.G., Gorman, G.S., Trenell, M.I., McFarland, R., Taylor, R.W., Turnbull, D.M., MacGowan, G.A., Blamire, A.M., Chinnery, P.F., 2012. Cardiomyopathy is common in patients with the mitochondrial DNA m3243A>G mutation and correlates with mutation load. Neuromuscul. Disord. 22, 592–596.
Kaufman, J.S., Dolman, L., Rushani, D., Cooper, R.S., 2015. The contribution of genomic research to explaining racial disparities in cardiovascular disease: a systematic review. Am. J. Epidemiol. 181, 464–472.
Khogali, S.S., Myosi, B.M., Beattie, J.M., McKenna, W.J., Watkins, H., Poulton, J.A., 2001. A common mitochondrial DNA variant associated with susceptibility to dilated cardiomyopathy in two different populations. Lancet 357, 1265–1267.
Lammertyn, L., Schutte, A., Schutte, R., 2011. Blood glucose and nocturnal blood pressure in African and Caucasian men: the SABPA study. Diabetes Res. Clin. Pract. 93, 235–242.
Li, B., Krishnan, V.G., Mort, M.E., Xin, F., Kamati, K.K., Cooper, D.N., Mooney, S.D., Radivojac, P., 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 25, 2744–2750.
Liu, C., Yang, Q., Hwang, S., Sun, F., Johnson, A.D., Shirihai, O.S., Vasan, R.S., Levy, D., Schwartz, F., 2012. Association of genetic variation in the mitochondrial genome with blood pressure and metabolic traits. Hypertension 60, 949–956.
Mahmood, S.S., Levy, D., Vasan, R.S., Wang, T.J., 2014. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 383, 999–1008.
Malan, L., Hamer, M., Frasure-Smith, N., Steyn, H.S., Malan, N.T., 2015. Cohort profile: sympathetic activity and ambulatory blood pressure in Africans (SABPA) prospective cohort study. Int. J. Epidemiol. 44, 1814–1822.
Mancia, G., Fagard, R., Narkiewicz, K., Redón, J., Zanchetti, A., Böhm, M., Christiaens, T., Cifkova, R., De Backer, G., Dominiczak, A., Galderisi, M., Grobbee, D.E., Ruilope, L.M., 2013. 2013 ESH/ESC guidelines for the management of arterial hypertension: the task force for the management of arterial hypertension of the European society of hypertension (ESH) and of the European society of cardiology (ESC). J. Hypertens. 31, 1281–1357.
Manica, A., Amos, B., Balloux, F., Hanihara, T., 2007. The effect of ancient population bottlenecks on human phenotypic variation. Nature 448, 346–348.
Mensah, G.A., 2013. Descriptive epidemiology of cardiovascular risk factors and diabetes in Africa. Prog. Cardiovasc. Dis. 56, 240–250.
Moran, A., Forouzanfar, M., Sampsonc, U., Chugh, S., Feigine, V., Mensah, G.A., 2013. The epidemiology of cardiovascular diseases in Sub-Saharan Africa: the global burden of diseases, injuries and risk factors 2010 Study. Prog. Cardiovasc. Dis. 56, 234–239.
Okin, P.M., Kjeldsen, S.E., Dahlöf, B., Devereux, R.B., 2011. Racial differences in incident heart failure during antihypertensive therapy. Circ. Cardiovasc. Qual. Outcomes 4, 157–164.
Owolabi, M.O., Mensah, G.A., Kimmel, P.A., Adu, D., Ramsay, M., Waddy, S.P., Ovbiagele, B., Rabadan-Diehl, C., Rasooly, R., Akarolo-Anthony, S.A., Rotimi, C., 2014. Understanding the rise in cardiovascular diseases in Africa: harmonising H3Africa genomic epidemiological teams and tools. Cardiovasc. J. Afr. 25, 134–136.
Pereira, L., Soares, P., Radivoiac, P., Li, B., Samuels, D.C., 2011. Comparing phylogeny and the predicted pathogenicity of protein variations reveals equal purifying selection across the global human mtDNA diversity. Am. J. Hum. Genet. 88, 433–439.
Polzin, T., Daneschmand, S.V., 2003. On Steiner trees and minimum spanning trees in hypergraphs. Oper. Res. Lett. 31, 12–20.
Robinson, M.T., Fischel-Ghodsian, N., Fraser, H.S., Nicholson, G.D., Grim, C.M., Wilson, D.M., Wilson, T.W., Grim, C.E., 2004. Genetic influences on the increase in blood pressure with age in normotensice subjects in Barbados. Ethn. Dis. 14, 57–63.
Rosa, A., Brehem, A., 2011. African human mtDNA phylogeography at-a-glance. J. Anthropol. Sci. 89, 25–58.
Salas, A., Richards, M., De La Fe, T., Lareu, M.-V., Sobrino, B., Sánchez-Diz, P., Macaulay, V., Carracedo, Á., 2002. The making of the African mtDNA landscape. Am. J. Hum. Genet. 71, 1082–1111.
Salas, A., Elson, J.L., 2015. Mitochondrial DNA as a risk factor for false positives in case-control association studies. J. Genet. Genomics 42, 169–172.
Samuels, D.C., Carothers, A.D., Horton, R., Chinnery, P.F., 2006. The power to detect disease associated with mitochondrial DNA haplogroups. Am. J. Hum. Genet. 78, 713–720.
Schutte, A.E., Schutte, R., Huisman, H.W., van Rooyen, J.M., Fourie, C.M., Malan, N.T., Malan, L., Mels, C.M., Smith, W., Moss, S.J., Towers, W.G., Kruger, S.H., Wentzel-Viljoen, E., Vorster, H.H., Kruger, A., 2012. Are behavioural risk factors to be blamed for the conversion from optimal blood pressure to hypertensive status in Black South Africans? A 5-year prospective study. Int. J. Epidemiol. 14, 1114–1123.
Soares, P., Abrantes, D., Rito, T., Thomson, N., Radivojac, P., Li, B., Macaulay, V., Samuels, D.C., Pereira, L., 2013. Evaluating purifying selection in the mitochondrial DNA of various mammalian species. PLoS One 8, e58993.
Taylor, R.W., Giordano, C., Davidson, M.M., d'Amati, G., Bain, H., Hayes, C.M., Leonard, H., Barron, M.J., Casali, C., Santorelli, F.M., Hirano, M., Lightowlers, R.N., DiMauro, S., Turnbull, D.M., 2003. A homoplasmic mitochondrial transfer ribonucleic acid mutation as a cause of maternally inherited hypertrophic cardiomyopathy. J. Am. Coll. Cardiol. 41, 1786–1796.
Thusberg, J., Olatubosun, A., Vihinen, M., 2011. Performance of mutation pathogenicity prediction methods on missense variants. Hum. Mutat. 32, 358–368.
Tuppen, H.A., Blakely, E.L., Turnbull, D.M., Taylor, R.W., 2010. Mitochondrial DNA mutations and human disease. Biochim. Biophys. Acta 1797, 113–128.
van der Walt, E.M., Smuts, I., Taylor, R.W., Elson, J.L., Turnbull, D.M., Louw, R., van der Westhuizen, F.H., 2012. Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease. J. Hum. Genet. 20, 650–656.
van der Westhuizen, F.H., Sinxadi, P.Z., Dandara, C., Smuts, I., Riordan, G., Meldau, S., Malik, A.N., Sweeney, M.G., Tsai, Y., Towers, G.W., Louw, R., Gorman, G.S., Payne, B.A., Soodyall, H., Pepper, M.S., Elson, J.L., 2015. Understanding the implications of mitochondrial DNA variation in the health of Black Southern African populations: the 2014 Workshop. Hum. Mutat. 36, 569–571.
van Oven, M., Kayser, M., 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. 30, E386–E394.
Wallace, D., 2010. Bioenergetics and the epigenome: interface between the environment and genes in common diseases. Dev. Disabil. Res. Rev. 16, 114–116.
Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., Schönherr, S., 2016. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. 44 (W1), W58–W63.
World Medical Association, 2004. World Medical Association declaration of Helsinki: ethical principles for medical research involving human subjects. J. Int. Bioethique 1, 124.
Yarham, J.W., Elson, J.L., Blakely, E.L., McFarland, R., Taylor, R.W., 2010. Mitochondrial tRNA mutations and disease. Wiley Interdiscip. Rev. RNA 1, 304–324.
Yu, X., Koczan, D., Sulonen, A.-M., Akkad, D.A., Kroner, A., Comabella, M., Costa, G., Corongiu, D., Goertsches, R., Camina-Tato, M., Thiesen, H.-J., Nyl, H.I., 2008. mtDNA nt13708A variant increases the risk of multiple sclerosis. PLoS One 3, e153.

Chapter 6: Mitochondrial DNA variation in oxidative stress and inflammation: the SABPA study

This chapter is presented in article format, as this manuscript is being prepared for submission to a peer reviewed scientific journal.

6.1 TITLE

Mitochondrial DNA variation in oxidative stress and inflammation: the SABPA study.

6.2 AUTHORS AND AFFILIATIONS

Marianne Venter[1], Leone Malan[2], Etresia van Dyk[1], Joanna L. Elson[1, 3], Francois H. van der Westhuizen[1]

1. Human Metabolomics, North-West University, Potchefstroom, South Africa
2. Hypertension in Africa Research Team (HART), North-West University, Potchefstroom, South Africa
3. Institute of Genetic Medicine, Newcastle University, United Kingdom

6.3 ABSTRACT

Mitochondrial DNA (mtDNA) variation has been implicated in several common complex and degenerative diseases, including cardiovascular disease. A common feature in many of these diseases is an inflammatory state. MtDNA variation can contribute to inflammation in several ways. We previously found no relationship between non-synonymous mtDNA variants and hypertension or hyperglycaemia in a bi-ethnic cohort. Here, were investigated the role of these variants in seven indicators of oxidative stress and inflammation, to determine if mtDNA are involved in the mechanisms leading up to disease phenotypes in this cohort. We found no significant relationships between non-synonymous mtDNA variants and the seven parameters investigated here, indicating that these variants are unlikely to impact on disease in this cohort, to an appreciable or measurable extent.

6.4 INTRODUCTION

Cardiovascular disease (CVD) as the global leading cause of morbidity and mortality (Mensah G A, 2013) has been the focus of a vast number of research studies in the past few decades (Hamer et al., 2015; Omboni et al., 2016). The goal has been to elucidate the underlying mechanisms and risk factors involved in the development and progress of CVDs which, together with recognised environmental and lifestyle factors (Malan & Malan, 2016), could also include genetic factors. Mitochondria play a central role in cell homeostasis, proliferation and apoptosis (Liu et al., 1996) most notably by producing the majority of cellular ATP via oxidative phosphorylation (OXPHOS), and being a major contributor of reactive oxygen species (ROS) and therefore oxidative stress (Nunnari & Suomalainen, 2012). ROS has been implicated in CVD phenotypes, such as hypertension, diabetes and atherosclerosis (de Champlain et al., 2004; Harrison et al., 2007), and is thought to contribute to the inflammatory state that is often observed in these phenotypes. As the major producer of ROS, mitochondrial dysfunction could also contribute to the aetiology of these diseases. Independent of ROS, mitochondrial dysfunction and the consequent decrease in ATP production could also directly contribute to inflammation, through increased apoptosis of i) vascular smooth muscle cells (VSMC), which decreases vascular integrity or, ii) monocytes, resulting in the release of cytokines (Yu et al., 2013; Yu & Bennett, 2016). The release of mtDNA itself into the cytosol and circulation (Nakahira et al., 2011; Oka et al., 2012; Shimada et al., 2012; West et al., 2015) has also been shown to result in the increased secretion of various pro-inflammatory factors (Sandhir et al., 2016). It therefore seems reasonable to consider the role of mtDNA variation when CVD phenotypes are investigated. MtDNA is a small (16 569 bp), circular DNA molecule located within the mitochondrial matrix. It encodes for two rRNAs, 22 tRNAs, and 13 polypeptides with catalytic importance for enzyme complexes I, III, IV and V of the OXPHOS system. MtDNA has a high mutation rate (Tuppen et al., 2010) and is maternally inherited, which enables us to recognise distinct lineages called haplogroups (Elson & Lightowlers, 2006). MtDNA changes can be classified as either rare variants, which are more likely to be pathogenic, or common population variants, which are often haplogroup-defining (Elson et al., 2006; Pereira et al., 2011).

We previously reported no relationship between mtDNA variation and two CVD phenotypes, namely hypertension and hyperglycaemia, in the bi-ethnic Sympathetic Activity and Ambulatory Blood Pressure in Africans (SABPA) cohort using the MutPred adjusted load hypothesis approach (Venter et al., 2017). However, the sympathetic nervous system is involved in the progression of hypertension and diabetes (Malan et al., 2013a; 2013b), which might out-weigh the influence of mtDNA variation, if any exists. To see whether mtDNA variation may impact on factors involved in CVD development, but that are more directly associated with mitochondrial (dys)function, we analysed the relationship between seven indicators of oxidative stress and inflammation, and mtDNA variation, using the MutPred adjusted load hypothesis approach (Venter et al., 2017). For an indication of oxidative stress status, we

used measurements of serum ROS, nitric oxide (NO), 8-hydroxy-deoxyguanosine (8-OHdG), thiobarbituric acid reactive substances (TBARS), and whole blood reduced glutathione (GSH). For inflammation, we used measurements of serum C-reactive protein (CRP), and tumor necrosis factor alpha (TNFα).

6.5 METHODS

6.5.1 Cohort description

Our sub-study is nested within the SABPA prospective cohort study. Details on participant recruitment and clinical assessments during the data collection phase have been described elsewhere (Malan et al., 2015). Inclusion criteria were urban Black African and Caucasian male and female teachers from South Africa, with similar socio-economic status, aged 20-65 years, from the North-West Province. Participants for the current investigation were enrolled in the project in 2008-2009 for baseline assessments. During recruitment and before consent, study participants were informed about the objectives and procedures of the study, including investigations of genes involved in CVD. The study complied with criteria on human research set by the Helsinki declaration (2004) and was approved by the Ethics Committee of the North-West University (NWU 00036-07-S6).

6.5.2 Sequencing and data management

DNA for sequencing was extracted from whole blood samples. Sequencing of the entire mtDNA molecule was done using the Ion Torrent PGM, as described elsewhere (van der Walt et al., 2012; Venter et al., 2017) for 363 participants. Variants were called using the revised Cambridge reference sequence (rCRS). MtDNA haplogroups were assigned according to the mtDNA variant list of each participant, using Haplogrep 2.0 (Weissensteiner et al., 2016). MutPred adjusted mutational loads were calculated as described by Venter et al. (2017). Briefly, using the MutPred system to assign a pathogenicity score between 0 and 1 to each variant within an individual's mtDNA, a mutational load can be calculated by adding pathogenicity scores of variants occurring in the same person. Because i) the rCRS is a European sequence; and ii) there exists genetic diversity between population groups, especially those of African origin, the number of mtDNA variants per person can differ significantly (Cavadas et al., 2015; Gurdasani et al., 2015). To minimise the influence of this population stratification on mutational loads, the loads were adjusted in two ways: firstly, only variants with a MutPred pathogenicity score above 0.5 were used, as the likelihood of pathogenicity of such variants can be considered an "actionable hypothesis". Most common population specific variants are excluded in this way. Secondly, MutPred mutational loads were divided by the number of variants used to calculate

mutational loads; this adjusts for the position of sequences from different lineages within the phylogeny, to that of the rCRS.

6.5.3 Biochemical parameters of oxidative stress status and inflammation

Seven indicators of oxidative stress and inflammation were used in this study: ROS was measured on a Bio-Tek FL600 microplate fluorescence reader (Bio-Tek, Instruments) in serum as peroxide content using a colorimetric assay as described by Hiyashi (2007); 8-OHdG was measured in urine using an LC-MS, and normalised against urine creatinine concentrations (Hu et al., 2004; Saude et al., 2007); NO bioavailability was indirectly estimated as the sum of plasma nitrite and reduced nitrate (NOx). NOx was analysed on a Universal ELX800 plate reader and GEN5 software (BioTek Instruments) applying an R&D Systems Inc kit (Schwedhelm et al., 2005). Inter- and intra-assay variability were 8.2 - 13.5% and 5.5%, respectively (Reimann et al., 2013). TBARS was measured in urine using an LC-MS methodology, and normalised against urine creatinine concentrations (Mokhaneli et al., 2016); reduced glutathione (GSH) was measured in whole blood using a Bioxytech GSH/GSSG-412 kit, on a Bio-Tek FL600 microplate fluorescence reader (Mels et al., 2014); ultra-high sensitivity CRP was measured in serum using the turbidimetric method with unicel DXC 800 (Huisman et al., 2012); high sensitivity TNF-α was analysed with the Quantikine High Sensitivity Human Tumor TNF-α enzyme-linked immunosorbent assay (HS ELISA; R&D Systems). The inter- and intra-assay variability were 15% and 17.8%, respectively (Jansen van Vuren et al., 2016b).

6.5.4 Lifestyle factors

Anthropometric measurements were performed in triplicate by level II anthropotometrists according to standardized procedures. Intra- and inter-observer variability was less than 10%. Body surface area (BSA) in m^2, was calculated according to the Mosteller formula (Mosteller, 1987). Total energy expenditure (TEE) or physical activity in 24 hours (kcal/day) considered resting metabolic rate (Actical®, Mini Mitter, Montreal, Quebec). Tobacco use, as indicated by serum cotinine (ng/ml), was measured using a homogeneous immunoassay on Modular ROCHE. Alcohol use, as indicated by the liver enzyme serum gamma glutamyl tranferase (cGGT) in U/L, was measured using the enzyme rate method on a Unicel DXC 800 (Jansen van Vuren et al., 2016a).

6.5.5 Statistical analyses

All statistical analyses were carried out using IBM SPSS Statistics (version 23). Within each gender/mtDNA background group, an *a priori* adjustment, as proposed by (Piepoli et al., 2016), was made to all biochemical parameters, for the following life-style confounders: age, BSA, TEE, serum

cotinine and cGGT. In addition to these five confounders, indicators of oxidative stress (ROS, 8-OHdG and TBARS) were also adjusted for serum CRP measurements. This was done by using an ANCOVA model, with the fixed and random factor windows left blank. The resulting adjusted values (residuals) for each biochemical parameter (indicators of oxidative stress and inflammation), were saved as new variables and used in further statistical analyses. For each parameter, means of the new *a priori* adjusted residuals or values were compared between participants that have variants with MutPred scores above 0.5, and participants that have no variants with MutPred scores above 0.5, using independent T-tests. Pearson's correlation analyses were used to assess the relationship between MutPred adjusted loads and the seven new *a priori* adjusted parameters investigated here. Also, A Bonferroni correction for multiple comparisons was applied throughout all statistical analyses.

6.6 RESULTS

Participants were divided into four groups according to gender (male and female) and African or European mtDNA background (macro-haplogroups L or MN respectively), and analysed separately as significant phenotypical discrepancies between these groups exist (Malan et al., 2015). Within the four gender/background groups, participants were further grouped into two categories: those with high MutPred-scoring variants (scores above 0.5), and those without. After making *a priori* adjustments as described under *Statistical Analyses,* T-tests were carried out to compare the means of *a priori* adjusted measurements for each marker tested, between these two groups. As can be seen in Table 6.1, no significant differences were found for any of the oxidative stress and inflammation markers tested. Pearson's correlation analyses were performed for MutPred adjusted loads against all *a priori* adjusted markers. As can be seen in Table 6.2, there were no significant correlations between MutPred adjusted loads and any of the *a priori* adjusted measurements for oxidative stress and inflammation markers, in any of the groups. While a correlation between MutPred adjusted loads and adjusted 8-OHdG measurements was borderline significant (r = 0.232; P-value = 0.03) in haplogroup MN females, this result was not robust against a correction for multiple testing.

6.7 DISCUSSION

Inflammation is frequently present in CVDs and is thought to be an important factor in the aetiology of disease phenotypes like atherosclerosis and hypertension. Because mitochondrial dysfunction has been shown to contribute to inflammation, mtDNA variation is a plausible genetic risk factor to investigate in such conditions. Here we examined a putative role for mtDNA variation in seven factors thought to be susceptible to changes as a result of mitochondrial dysfunction.

Table 6.1: Comparing means of oxidative stress and inflammation markers of participants who have high MutPred-scoring mtDNA variants, with those who do not

	Variants with MutPred scores above 0.5 present	Total cohort ($N = 363$)		Haplogroup L males ($N = 100$)		Haplogroup L females ($N = 94$)		Haplogroup MN males ($N = 75$)		Haplogroup MN females ($N = 94$)	
		Mean ± SEM	*t* (df)	Mean ± SEM	*t* (df)	Mean ± SEM	*t* (df)	Mean ± SEM	*t* (df)	Mean ± SEM	*t* (df)
Serum ROS (mg/L)	No	95.02 ± 1.66	0.79 (355)	84.07 ± 2.09	-0.17 (97)	110.17 ± 3.26	1.52 (88)	77.95 ± 2.25	1.55 (73)	102.09 ± 3.3	-0.44 (91)
	Yes	92.06 ± 3.15		84.23 ± 4.33		102.62 ± 6.44		73.57 ± 3.11		108.18 ± 6.85	
8-OHdG (ng/mg creatinine)	No	31.7 ± 3.63	-0.35 (339)	22.77 ± 4.3	-0.96 (95)	26.74 ± 3.86	-0.81 (85)	61.29 ± 18.37	1.24 (67)	29.47 ± 3.01	-2.24 (86)
	Yes	38.95 ± 4.16		34.52 ± 10.09		32.61 ± 6.2		37.59 ± 8.27		45.52 ± 7.24	
TBARS (mg/g creatinine)	No	0.13 ± 0.01	-1.42 (313)	0.15 ± 0.02	-0.91 (87)	0.16 ± 0.02	-1.06 (74)	0.09 ± 0.01	-0.58 (67)	0.11 ± 0.01	-0.66 (79)
	Yes	0.15 ± 0.02		0.24 ± 0.07		0.19 ± 0.04		0.1 ± 0.01		0.12 ± 0.01	
Serum CRP (mg/L)	No	6.32 ± 0.55	-1.38 (358)	5.27 ± 0.92	-0.39 (98)	11.62 ± 1.27	-1.54 (89)	2.35 ± 0.35	0.18 (73)	3.27 ± 0.56	-0.92 (92)
	Yes	5.42 ± 0.79		5.11 ± 1.87		14.58 ± 3.25		2.4 ± 0.36		4.33 ± 0.95	
Serum TNFα (IU/ml)	No	0.23 ± 0.02	0.05 (358)	0.32 ± 0.04	0.36 (98)	0.24 ± 0.02	0.48 (89)	0.18 ± 0.02	-1.19 (73)	0.12 ± 0.02	0.13 (92)
	Yes	0.2 ± 0.02		0.32 ± 0.04		0.21 ± 0.05		0.22 ± 0.03		0.11 ± 0.02	
Total serum NO (µmol/L)	No	8.94 ± 0.99	1.13 (356)	10.6 ± 1.49	1.18 (98)	12.96 ± 1.48	0.4 (89)	3.13 ± 1.16	0.15 (72)	5.17 ± 3.24	0.59 (91)
	Yes	5.35 ± 0.67		8.75 ± 1.7		11.94 ± 2.01		2.78 ± 0.92		2.62 ± 0.64	
Whole blood GSH (µM)	No	876.4 ± 13.77	0.59 (345)	934.2 ± 20.52	0.69 (98)	855.7 ± 20.69	-0.58 (88)	861.9 ± 27.53	-0.41 (73)	828.2 ± 46	0.97 (80)
	Yes	845.5 ± 22.22		899.5 ± 49.86		893.5 ± 38.41		866.3 ± 33.52		767.8 ± 45.38	

Independent T-tests were carried out to compare the means of oxidative stress and inflammation markers, with prior *a priori* confounder adjustments, of participants who carry high MutPred-scoring mtDNA variants, with those who do not. Means and standard deviations (SEM). The t statistic (*t*) and degrees of freedom (df) are presented using prior adjusted new residual values of markers. No statistically significant differences were found. ROS: reactive oxygen species; 8-OHdG: 8-hydroxy-deoxyguanosine; TBARS: thiobarbituric acid reactive substances; CRP: C-reactive protein; TNFα: tumor necrosis factor alpha; NO: nitric oxide; GSH: total glutathion; MutPred: a pathogenicity score assigned to a variant using the MutPred program (*mutpred.mutdb.org/about.html)*.

Table 6.2: Pearson's correlations comparing oxidative stress and inflammation markers, with MutPred adjusted loads

	Total cohort (*N* = 363)		Haplogroup L males (*N* = 100)		Haplogroup L females (*N* = 94)		Haplogroup MN males (*N* = 75)		Haplogroup MN females (*N* = 94)	
	Mean ± SD	*r*	Mean ± SD	*r*	Mean ± SD	*r*	Mean ± SD	*r*	Mean ± SD	*r*
Serum ROS (mg/L)	93.43 ± 28.25	.043	84.35 ± 18.92	-.026	107.7 ± 27.94	.178	76.2 ± 15.95	.185	103.8 ± 33.32	-.053
8-OHdG (ng/mg creatinine)	33.96 ± 50.9	-.023	25.88 ± 40.13	-.102	27.37 ± 30.98	-.094	50.05 ± 84.84	.142	35.74 ± 34.18	-.232
TBARS (mg/g creatinine)	0.14 ± 0.14	-.089	0.17 ± 0.22	-.115	0.16 ± 0.14	-.148	0.1 ± 0.06	-.062	0.12 ± 0.06	-.068
Serum CRP (mg/L)	5.91 ± 8.48	-.075	5.19 ± 8.23	-.040	11.95 ± 11.49	-.170	2.46 ± 2.2	.025	3.67 ± 4.98	-.092
Serum TNFα (IU/ml)	0.22 ± 0.23	-.002	0.32 ± 0.31	.026	0.24 ± 0.21	.045	0.2 ± 0.17	-.129	0.12 ± 0.16	.013
Total serum NO (µmol/L)	7.6 ± 13.32	.064	10.12 ± 11.86	.147	12.78 ± 12.31	.045	2.74 ± 6.17	.010	3.92 ± 17.31	.056
Whole blood GSH (µM)	864.2 ± 219.5	.027	925.5 ± 197.2	.073	863.4 ± 175.6	-.066	859.1 ± 187.6	-.056	798.1 ± 290.4	.096

Pearson's correlations were carried out to investigate the relationship between prior *a priori* adjusted measurements of oxidative stress and inflammation markers, and MutPred adjusted loads. Means and standard deviations (SD) as well as Pearson's correlation coefficients (*r*) of oxidative stress and inflammation markers are presented. ROS: reactive oxygen species; 8-OHdG: 8-hydroxy-deoxyguanosine; TBARS: thiobarbituric acid reactive substances; CRP: C-reactive protein; TNFα: tumor necrosis factor alpha; NO: nitric oxide; GSH: total glutathion; MutPred: a pathogenicity score assigned to a variant using the MutPred program (*mutpred.mutdb.org/about.html)*.

ROS is produced in many pathways, most notably by NADPH oxidase and enzyme complexes of the electron transport chain (ETC), and is an indicator of oxidative stress (Harrison et al., 2007). ROS levels have previously been shown to be positively associated with blood pressure in Black South African males in the current cohort (Kruger et al., 2012), although in a later publication on the SABPA cohort, the significance of this relationship was lost when HIV+ participants were excluded (Mels et al., 2014). In the current study, no association was found between mtDNA variation and *a priori* adjusted ROS measurements in any of the groups. The bioavailability of endothelial derived NO, which is important for vasodilation, can be reduced by ROS via quenching, leading to hypertension (Harrison et al., 2007). Estimated NO bio-availability was previously shown to be higher in Black South African participants than in their Caucasian counterparts, in the SABPA cohort (Reimann et al., 2013). We did not find a relationship between MutPred adjusted loads and *a priori* adjusted NO measurements in any of the gender/mtDNA background groups.

The measurement of ROS levels in serum is not direct, but is rather derived from measurements of reactive oxygen metabolites, which makes this measurement susceptible to inaccuracies. Therefore, the inclusion of other indicators of oxidative stress can be useful. 8-OHdG is a urinary bio-marker for oxidative damage to DNA, while TBARS are formed during lipid peroxidation (Janero, 1990). High levels of oxidative molecules cause wide-spread damage to proteins, lipids and nucleic acid. When this damage is repaired in DNA, 8-OHdG is cleaved and excreted in the urine. Elevated levels of 8-OHdG have been associated with atherosclerosis and diabetes (Wu et al., 2004). During lipid peroxidation malondialdehyde is formed, which can be measured in urine as it reacts with thiobarbituric acid in a TBARS assay (Janero, 1990). In the SABPA cohort, TBARS were previously shown to be positively associated with arterial compliance and vascular resistance in Black South African participants (Mokhaneli et al., 2016). Also, 8-OHdG levels in urine were previously found to be inversely correlated with blood pressure in Black South African males, despite having higher ROS levels (Mels et al., 2014). Mels et al. (2014) suggested that this counter-intuitive relationship is indicative of mitochondrial hormesis, where oxidative stress in Black South African males of this cohort are at levels where protective mechanisms are upregulated, and not yet at inhibitory levels. This theory was further supported by demonstrating that total GSH activity, an indicator of anti-oxidant capacity, correlated well with ROS levels, showing that protective mechanisms are upregulated during increased oxidative stress. We however, could not find any association between MutPred adjusted loads and measurements for 8-OHdG, TBARS or GSH, in any of the groups. This might indicate that mtDNA variants in protein encoding genes, are not associated with the altered measurements of oxidative stress markers in this cohort. It is possible that other producers of ROS, such as NADPH oxidase, are contributing to the oxidative stress in these participants, to a greater extent than enzymes of the ETC (Harrison et al., 2007).

Indeed, inflammation is central to the proposed mechanisms that link mtDNA to CVDs. While it seems that in the SABPA cohort, mtDNA variation is not associated with oxidative stress, the possibility still exists that mtDNA variation could be implicated in inflammation, independent of ROS (Chapter 2). TNFα is a cytokine that is produced mainly by macrophages as part of the acute phase reaction of inflammation (Wajant et al., 2003). Synthesis of CRP is triggered by cytokines as a response to the onset of inflammation, within two hours. CRP facilitates phagocytosis of dead or dying cells by binding to surface proteins (Pepys & Hirschfield, 2003). Therefore, high sensitivity tests for both TNFα and CRP, make them useful markers for inflammation, whereby the risk for developing CVDs can be assessed. In the SABPA cohort, it was previously reported that TNFα and CRP levels were significantly higher in Black South African participants compared to their Caucasian counterparts, and in Black South African males, cardiac stress was significantly associated with TNFα (Jansen van Vuren et al., 2016a; 2016b). However, we could not find any relationships between *a priori* adjusted measurements for either of these markers, and MutPred adjusted loads, in any of the groups. Thus, as was found for

the markers of oxidative stress above, it seems unlikely that mtDNA variants within protein-encoding genes have an impact on markers for inflammation in the current cohort.

While at first glance it seems that mtDNA variation is not involved in the aetiology of CVDs in the SABPA cohort, it is also possible that any down-stream effect of the mtDNA variation reported here is compensated for by other regulatory pathways, or masked by the consequences of other more influential factors. This is especially plausible since the participants of the SABPA cohort were not selected based on clinical features, and so few participants were severely diseased. Many environmental and behavioural risk factors are known to significantly contribute to the development and progression of CVDs (Malan et al., 2008; 2015). Genetic factors such as mtDNA or nuclear DNA variation would not cause CVDs, but would rather impact on disease susceptibility or alter the course of disease (Achilli et al., 2011), likely to a much lesser extent. Therefore, if the presence of one or more high MutPred-scoring (mildly deleterious) mtDNA variants could lead to mitochondrial dysfunction that consequently impacts on CVDs, it is possible that disease progression is not advanced enough in the SABPA cohort to demonstrate this impact, or that the influence of environmental and life-style factors in this cohort (Malan et al., 2013b; van Deventer et al., 2015) greatly out-weighs that of mtDNA variants. Future studies should include more direct measurements of mitochondrial function in these participants, which would enable us to know if the mtDNA variants reported here lead to mitochondrial dysfunction that is still mild enough to be tolerated by SABPA participants, and thus is masked from the analyses. Another plausible line of investigation to follow in this and other similar cohorts is the role of mtDNA copy number in CVD risk factors. Studies have shown that high mtDNA copy number is associated with a lower prevalence of microalbuminuria (Lee et al., 2009), and might protect against the development of diabetes type 2 (Lee et al., 1998; Zhou et al., 2016), although this latter finding has been contested (Singh et al., 2007). MtDNA copy number in peripheral blood cells was also altered in diabetic nephropathy (Czajka et al., 2015) and diabetic retinopathy (Malik et al., 2015). Replication of these studies in different populations might help to elucidate the putative relationship between mtDNA variation and these disease phenotypes.

Chapter 7: Utilising transmitochondrial cytoplasmic hybrid cells to test the MutPred mutational load hypothesis

7.1 INTRODUCTION

In previous chapters, it was hypothesised that mildly deleterious mtDNA variants may play a role in human disease by altering mitochondrial function and impacting on several downstream processes. However, by utilising the MutPred adjusted load method in the SABPA cohort, no associations could be found between high MutPred scores/adjusted mutational loads and hypertension or diabetes levels (Venter et al., 2017), or any measured markers of oxidative stress and inflammation (Chapter 6). Thus, questions remain as to whether the differences between those with high MutPred adjusted loads and those with lower loads, are not pronounced enough to be of statistical significance in the SABPA data, or are masked by other more influential regulatory processes or risk factors, or whether genetic variation as defined by MutPred load indeed translates to a significant impact on mitochondrial function. This latter fundamental biochemical question lies at the core of this study and, as described in this chapter, was further investigated using transmitochondrial cytoplasmic hybrid (cybrid) cells which were available for this study. Cybrid cells are generated by first depleting the host cell line of its own mitochondrial DNA (mtDNA) to create rho0 cells. Donor mtDNA is then introduced to rho0 cells by fusion with donor blood platelets (Chomyn, 1996). In this way, several cell lines can be engineered that have the same nuclear DNA background (that of the host cell line), but unique mtDNA. In theory, any differences measured in mitochondrial function, such as mitochondrial respiration, can then be attributed to mtDNA variation between cybrid cell lines (assuming all other possible variables remain unaffected by the process). It is important to note here that this process isolates the donor mtDNA from all other donor cell elements, eliminating the influence of factors such as the gender and non-mitochondrial DNA disease phenotype of the donating person (Wilkins et al., 2014). Although chromosomal instability of rho0 cells is a cause for concern in the process (Singh et al., 2005), several studies have used this method to demonstrate the role of mtDNA variation in mitochondrial function or disease (reviewed in Wilkins et al., 2014). For example, studies have shown that a variant in the *MT-ND5* gene, m.13565C>T, resulted in reduced mitochondrial membrane potential and defective mitochondrial function (McKenzie et al., 2007), which was associated with decreased calcium uptake by ND5 mutant cybrids (McKenzie et al., 2016). In a study by Silva et al. (2013), cybrids generated using blood platelets from Alzheimer Disease (AD) and mild cognitive impairment (MCI) patients had reduced COX activity and mitochondrial respiration when compared to control cybrids. However, conflicting results have previously been reported for both AD and Parkinson Disease cybrid studies

(Howell et al., 2005). In a series of studies, Kenney et al. (2014a, 2014b, 2016) showed that cybrids cell lines with different haplogroups (L, H or J) had significant alterations in bio-energetic profiles, reactive oxygen species (ROS) production, as well as in the expression of several mitochondrially encoded protein genes and nuclear encoded complement, inflammation, apoptosis and Wnt signalling pathway, and diabetes-related genes.

In this study, cybrid cells were used to investigate whether mtDNA variants with high MutPred scores, which are more likely to be mildly deleterious mtDNA variants (Li et al., 2009), influenced mitochondrial respiration. As blood samples from the SABPA study cohort were not available for this investigation, cybrid cells were prepared from samples available from a previous, unrelated study, and were selected from patients and controls according to specific mtDNA variant-based selection criteria. Mitochondrial respiration in these cybrids was then measured using a Seahorse XF^{e} 96 extra cellular flux analyser. As mentioned earlier, the process by which cybrids cell lines are produced eliminates the influence of non-mtDNA related characteristics of the donor cells. It is therefore not essential to use cybrid cell lines produced from SABPA participants to investigate the fundamental biochemical question stated above.

7.2 METHODS

7.2.1 Ethics approvals

The following ethics approvals from the Health Sciences Ethics Office for Research at the North-West University, Potchefstroom Campus, applies to the work presented here. For DNA sequencing of the samples described here, which investigated the aetiology of myalgic encephalomyelitis (also known as chronic fatigue syndrome or CFS), approval according to application number NWU 00 102-12-S1 applies. For the development and evaluation of cybrids the application NWU-00358-16-S1 applies.

7.2.2 Sequencing and mtDNA variation analyses

DNA extraction, sequencing and mtDNA variant calling was done as previously described (Chapter 3; van der Walt et al., 2012; Venter et al., 2017). Only consensus variants were used. Haplogroups were assigned using Haplogrep 2.0 (*haplogrep.uibk.ac.at*) and MutPred scores were calculated for all non-synonymous variants using the MutPred program (*mutpred.mutdb.org*). MutPred total loads were calculated by summing the MutPred scores of all non-synonymous variants for a cell line. MutPred

adjustment 1 loads[2] were calculated by summing MutPred scores of only non-synonymous variants with MutPred scores above 0.5 (mildly deleterious variants). Conservation and GenBank frequency details for each variant were obtained using MITOMAP. Variants were identified as haplogroup markers using Build 17 of Phylotree (*http://www.phylotree.org*). Phylogenetic median joining networks (Bandelt et al., 1999) were produced using the NETWORK (version 5) and NETWORK Publisher (version 2.0.0.1) software packages (*http://www.fluxus-engineering.com/sharenet.htm*).

7.2.3 Cybrid cell line selection and development

Previous repeatability tests done at the Mitochondrial Research laboratory (van Dyk, 2016) on the Seahorse XF^e 96 analyser (Agilent Technologies) have shown that inter-assay variability (between different assay plates) can be up to 22% (data not shown). This is comparable to the inter-assay variability previously reported by Seahorse Bioscience and others for the Seahorse XF^{e}24 analyser (Dranka et al., 2011). Consequently, the goal was to include all samples to be compared in a single analysis run. To minimise intra-assay variation and allow enough repeats of each sample (10-12 wells per sample), the number of cell lines that could be analysed on one 96-well plate was limited to eight. From the limited samples available, eight were selected according to specific criteria that best defined several MutPred load/mtDNA variant groups. The properties of these samples were as follows (see also Tables 7.1 and 7.2):

- All samples were of comparable haplogroups (U and R)
- three samples had no variants with MutPred scores above 0.5
- three samples each had one variant with a MutPred score above 0.5
- and two samples had multiple variants with MutPred scores above 0.5

Previously, blood was obtained from patients and controls using tubes containing EDTA (7.2 mg) and the blood platelets were then isolated and frozen using the method described by Chomyn (1996). As a gift from Dr. Leo Nijmants (Radboud Nijmegen University Medical Centre, Nijmegen, The Netherlands), 143B osteosarcoma rho0 cells were cultured in DMEM, supplemented with 10% (v/v) FBS, 1x penicillin-streptomycin (pen-strep), 100 µg/mL 5-bromo-2'-deoxyuridine, and 50 µg/mL uridine. Cybrids were then produced by polyethylene glycol fusion of patient and control blood platelets with 143B rho0 cells as described by Chomyn (1996). The newly formed cybrid cell lines were then allowed to grow and divide for 35 passages in the same media as above, but with 5-bromo-2'-deoxyuridine now excluded.

[2] Note that the "MutPred adjustment 1 load" differs from the "MutPred adjusted load" described in previous chapters. Here, only the first adjustment step was applied (namely the exclusion of variants with MutPred scores < 0.5), but not the second step (adjusting for the sequence position in the phylogeny), as all cybrid cell lines were of comparable haplogroups.

7.2.4 Bio-energetic flux analysis

Bio-energetic flux measurements of cybrid cell lines were performed using the Seahorse XF^e 96 analyser according to optimized conditions for cybrid cells (van Dyk, 2016). One day before the assay commenced, cells for all cell lines were seeded at 13 000 cells per well into two 96-well microtiter Seahorse cell culture plates (Agilent Technologies, CAT# 102601-100) and incubated for 24 hours in either glucose [DMEM (from Gibco, CAT# 41966052) containing 1 mM sodium pyruvate and 4.5 g/L glucose, supplemented with 10% (v/v) FBS and 1x pen-strep] or galactose [DMEM (from Sigma, CAT# D5030) supplemented with 10 mM galactose, and 10% (v/v) FBS and 1x pen-strep] rich growth medium. Each cell lines had 10 to 12 repeats per plate, and the four corner wells were used as blank controls. One hour before the assay commenced, cells were washed with assay medium (containing 1 mM sodium pyruvate and 5 mM glucose or 10 mM galactose), and the plates were placed in a CO_2-free incubator. A Mito stress test kit (from Agilent technologies, CAT # 103015-100) was used according to the manufacturer's directions, to calculate several parameters of mitochondrial OXPHOS by measuring oxygen consumption rates (OCR). The following conditions, as optimised previously for 143B cybrids by van Dyk (2016), were used: at the onset of the Mito stress test, in each well, basal respiration is measured in assay medium (containing either glucose or galactose as described above); oligomycin with an end concentration of 1 µM is first injected to inhibit ATP-synthase; next FCCP with an end concentration of 0.5 µM is injected to uncouple OXPHOS; finally rotenone and antimycin A with end concentrations of 0.5 µM are injected to completely stop mitochondrial respiration. To account for any differences in cell seeding growth density, all plates were normalized to the DNA content per well. This was done using the CyQUANT Cell Proliferation Assay kit (Thermo Fisher Scientific, CAT# C7026), which uses a proprietary green fluorescent CyQUANT® GR dye. When this dye binds to cellular nucleic acids, it exhibits a fluorescence enhancement which is measured using a Synergy HT microplate reader (from BioTek) with excitation at 485 ± 10 nm and emission detection at 530 ± 12.5 nm. The normalised OCR values were calculated with the Seahorse Wave software (Version 2.3) using the normalisation function expressed per fluorescent arbitrary unit (FAU).

7.2.5 mtDNA copy number determination

DNA was isolated from cybrid cells obtained at the various time points (passage numbers 17, 22, 26) and at the time of cell seeding (passage number 35). The relative mtDNA copy number (RMCN) was determined using real-time PCR as previously described (Meissner-Roloff, 2009) and optimised (van Dyk, 2013). The TaqMan® Gene Expression Assays kit (Thermo Fisher Scientific, CAT# 4331182) was used, which included primers for the nuclear β-globin gene (Hs00758889_s1) and the mitochondrial ND2 gene (Hs02596874_g1). Both of these genes were measured in triplicate in each

sample. RMCN was expressed as the number of mtDNA copies per two copies of β-globin gene (thus per cell, as each cell contains two copies of the β-globin gene).

7.2.6 Statistical analyses

All statistical analyses were carried out using IBM SPSS Statistics (version 23) or Prism GraphPad software (version 6.05). A Fisher's exact test was used to see whether variants appearing only once in Phylotree were more likely to have high MutPred scores or not (Pereira et al., 2011). Pearson's correlation tests were used to see whether different bio-energetic parameters correlated with basal respiration rates. Pearson's correlations were also used to test the relationship between basal respiration rates and several genetic parameters. R^2-values were calculated by squaring Pearson's correlation coefficients (r). Bonferroni corrections for multiple testing were applied throughout all statistical analyses, where required. To test the significance of differences between dependant correlation coefficients, a t-test was done (Chen & Popovich, 2002). To assess the differences in bio-energetic parameter measurements either between individual cell lines or mtDNA variant groups, one-way analyses of variance (ANOVA) with post-hoc Tukey tests were used.

7.3 RESULTS

To see whether the culmination of high MutPred-scoring mtDNA variants have an influence on mitochondrial respiration, as a measurement of mitochondrial function, a cybrid cell line model was utilised. Eight different cybrid cell lines previously created using 143 osteosarcoma rho0 cells as base cell line, and donor mtDNA from patients and controls in a *myalgic encephalomyelitis* cohort were selected, according to the criteria listed in *Section 7.2*. As mentioned, the process by which cybrid cells are produced eliminates the influence of the non-mtDNA characteristics associated with the donor disease phenotype. The genetic properties of the selected cybrid cell lines are given in Tables 7.1 and 7.2. All cybrid cell lines had comparable haplogroups (U2, U4, U5 and R1) and either had zero, one or more than two mtDNA variants with MutPred scores above the "actionable hypothesis" threshold (0.5). Cell lines having no variants with MutPred scores above 0.5 can also be described as having low MutPred adjustment 1 loads, while those with one or more variants with MutPred scores above 0.5 will have higher MutPred adjustment 1 loads. In Table 7.1, a detailed description of all non-synonymous coding mtDNA, mt-tRNA and mt-rRNA variants is given for each cell line. Negative selection against deleterious variants over time (Elson et al., 2004) has resulted in less common variants being more likely to be mildly deleterious. As such, variants with higher MutPred scores (likely to be mildly deleterious) are also more likely to be rare (Pereira et al., 2011). In this test group, no individual variant with a MutPred score above 0.5 was present in more than one participant.

Table 7.1: Consensus mt-tRNA, mt-rRNA and non-synonymous mtDNA variants for all cybrid cell lines

Gene	AA change	GenBank freq. (%)	Conser-vation (%)	MutPred score	Phylo-tree freq.	Nearest HG on Phylotree	SA1 (U2e1)	SA5 (U5b2a5)	SA19 (U4c2a)	SA6 (U5b1c2)	HC2 (U5a1c1a)	SA11 (U5b2a1b)	HC3 (U5b2b3a)	SA16 (R1a1a2)
MT-CYB	T-I	24091 (75.15)	49	0.17	4	HV	*14766T*	*14766T*	*14766T*	*14766T*	*14766T*	*14766T*	*14766T*	*14766T*
MT-ATP6	M-T	153 (0.48)	24	0.20	10	U4c			8705C					
MT-CYB	M-T	396 (1.24)	76	0.23	1	U4			15693C					
MT-COIII	V-I	1344 (4.19)	91	0.25	4	U5a		*9477A*		*9477A*	9477A	*9477A*	*9477A*	
MT-COIII	V-I	221 (0.69)	82	0.26	13	H1b	*9966A*							
MT-ND2	N-S	219 (0.68)	31	0.29	2	U5b						4732G		
MT-ATP8	M-I	0 (0.00)	20	0.32	0	-		**8491T***						
MT-CYB	T-A	642 (2.00)	80	0.37	5	U5a					15218G			
MT-ATP6	T-A	31527 (98.34)	71	0.37	4	R14	*8860G*	*8860G*	*8860G*	*8860G*	*8860G*	*8860G*	*8860G*	*8860G*
MT-ND5	I-V	91 (0.28)	98	0.38	3	U5b							12634G	
MT-ATP6	I-V	65 (0.20)	2	0.39	3	R1a								8887G
MT-ND4	F-L	74 (0.23)	13	0.39	2	U4c			10907C					
MT-CYB	T-A	31512 (98.29)	18	0.45	9	U4b	*15326G*	*15326G*	15326G	*15326G*	*15326G*	*15326G*	*15326G*	*15326G*
MT-CYB	H-R	741 (2.31)	51	0.46	5	U5a					14793G			
MT-ATP8	V-A	55 (0.17)	11	0.51	2	R1a								8388C
MT-ND5	T-A	62 (0.19)	24	0.56	1	U5b							13630G	
MT-ND5	T-M	24 (0.07)	60	0.60	2	U5a					**13802T**			
MT-ND5	Q-R	287 (0.90)	62	0.61	4	U5b							13637G	
MT-ND2	N-D	1528 (4.77)	91	0.63	6	R1a								4917G
MT-COI	T-A	0 (0.00)	96	0.63	0	-				**6444G**				
MT-ND1	V-A	62 (0.19)	24	0.65	3	U5b							3338C	
MT-ND2	V-I	59 (0.18)	93	0.66	3	N11b						*3736A*		
MT-ATP6	A-V	11 (0.03)	24	0.72	1	R1a								**8765T**
MT-ND5	P-S	50 (0.16)	89	0.72	1	R1a								13948T

Table 1 continued...

Gene	AA change	GenBank freq. (%)	Conser-vation (%)	Position in tRNA structure	Phylo-tree freq.	Nearest HG on Phylotree	SA1 (U2e1)	SA5 (U5b2a5)	SA19 (U4c2a)	SA6 (U5b1c2)	HC2 (U5a1c1a)	SA11 (U5b2a1b)	HC3 (U5b2b3a)	SA16 (R1a1a2)
MT-RNR1	rRNA	4118 (12.85)	40	-	59	U6	*709A*							
MT-RNR1	rRNA	31410 (97.98)	98	-	4	N10	*750G*	*750G*	*750G*	*750G*	*750G*	*750G*	*750G*	*750G*
MT-RNR1	rRNA	107 (0.33)	24	-	3	R1								1391C
MT-RNR1	rRNA	30179 (94.14)	87	-	14	P2	*1438G*	*1438G*	*1438G*	*1438G*	*1438G*	*1438G*	*1438G*	*1438G*
MT-RNR2	rRNA	225 (0.70)	58	-	1	U5b						1721T	1721T	
MT-RNR2	rRNA	2550 (7.95)	22	-	7	U2'4'	1811G		1811G					
MT-RNR2	rRNA	24784 (77.31)	84	-	9	U2	2706G	*2706G*	*2706G*	*2706G*	*2706G*	*2706G*	*2706G*	*2706G*
MT-RNR2	rRNA	184 (0.57)	56	-	4	U5b							2755G	
MT-RNR2	rRNA	10 (0.03)	100	-	2	M7	***3140G***							
MT-RNR2	rRNA	1350 (4.21)	36	-	4	U5a		*3197C*		*3197C*	3197C	3197C	*3197C*	
MT-TC	tRNA	57 (0.18)	0	Stem	2	R1a								5823G
MT-TD	tRNA	43 (0.13)	96	loop	1	R1a								7547C
MT-TL2	tRNA	4193 (13.08)	96	Top/stem	3	U5a	*12308G*	*12308G*	*12308G*	*12308G*	12308G	*12308G*	*12308G*	
MT-TT	tRNA	124 (0.39)	31	loop	3	U5b							15905C	
MT-TT	tRNA	229 (0.71)	60	loop	1	U2e	15907G							

The mtDNA variants found in the rRNA, tRNA and protein encoding genes are listed for each cybrid cell line used, with some descriptive data to the left. Transversions are marked with an asterisk. GenBank frequency was derived by MITOMAP from 32 069 sequences available on GenBank as of August 2016. Conservation is across 45 species. MutPred scores refer to pathogenicity scores assigned by the MutPred program for non-synonymous changes. Phylotree frequency refers to the number of times a variants is present as a haplogroup marker in the phylogenetic tree found at *www.phylotree.org*, Built 17. Where a variant listed as a haplogroup marker did not match the haplogroup of the cell line it appears in, the phylogenetically nearest haplogroup where the variant appears was given (indicated in italics). Very rare variants (GenBank frequency < 0.1%) are indicated in bold.

However, from Table 7.1 it can be seen that only three of these high-scoring variants were considered rare in the population (GenBank frequency < 0.1%, shown in bold). Of these, one variant (m.6444A>G) present in cell line SA6, was not present on GenBank at all, and has also not been reported according to MITOMAP. The same was true for one other variant in this cohort, m.8491A>T present in SA5, which is also the only non-synonymous transversion in this dataset. It should be noted that this variant has a MutPred score of only 0.32, well below the "actionable hypothesis" threshold. This would suggest that the variant on its own is not deleterious, despite being rare. It might, however, influence mitochondrial function against specific haplogroup backgrounds. At the very least, this demonstrates that even with more than 30 000 mtDNA sequences available on GenBank, not all existing variants have been documented, and more sequencing should be done.

As all the cybrid cell lines were of comparable haplogroups (U2, U4, U5 and R1), several variants were present in all or multiple cybrid cell lines of corresponding haplogroups. Figure 7.1 is a functional network, produced by using only non-synonymous variants. As can be seen, cybrids cell lines SA5, SA6, SA11, HC2 and HC3 (all haplogroup U5) cluster together on the left of the network, while SA1, SA16 and SA19 (haplogroups U2, R1 and U4) respectively clustered to the right of the network. To see whether variants were considered to be haplogroup markers, their presence on Phylotree (Build 17) was noted. Only the two very rare variants mentioned above (m.6444A>G and m.8491A>T) were not listed as haplogroup markers. In Table 7.1, the amount of times a variant appears as a haplogroup marker on Phylotree can be seen; very few variants were present only once. Although non-synonymous variants with MutPred scores above 0.5 were more likely than those with scores below 0.5 to be found only once on Phylotree, this difference was not significant (Fisher's $P = 0.19$). Not all variants shown to be haplogroup markers defined the haplogroup assigned to the corresponding cybrid cell line. In these cases, the phylogenetically nearest haplogroup was listed. These "out of place" variants are indicated in italics in Table 7.1, and were more common among mt-RNA genes than protein coding genes.

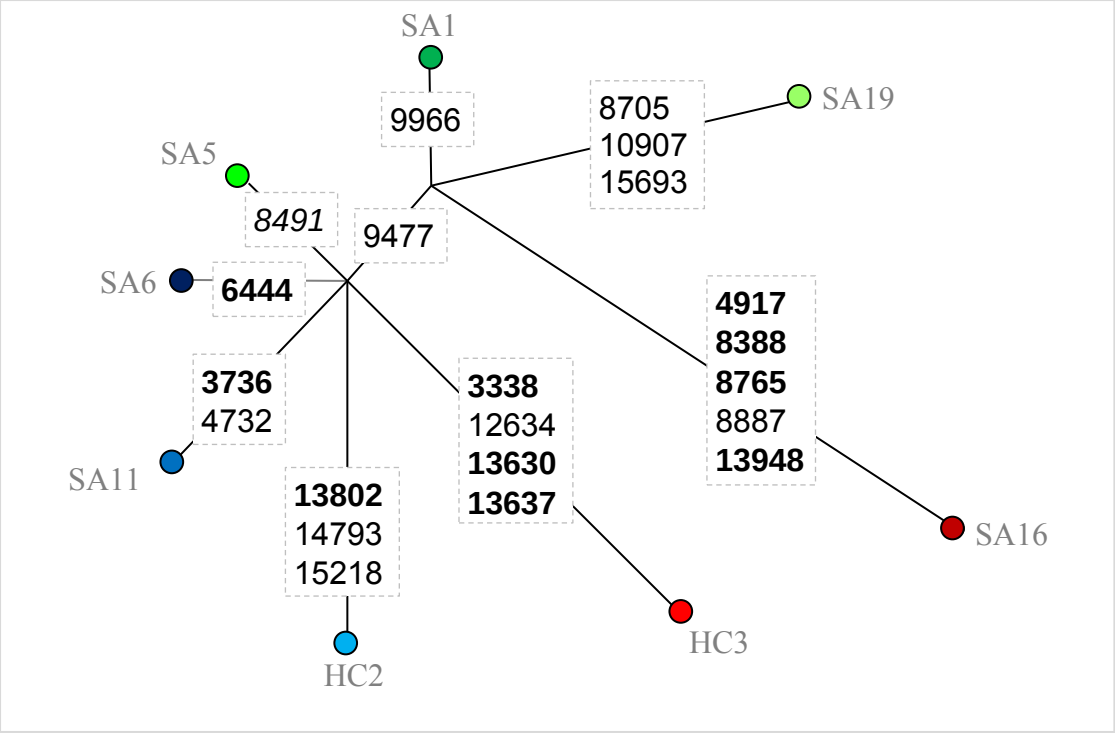


Figure 7.1: Functional network analysis of eight cybrid cell lines. A functional network showing non-synonymous variants for each cybrid cell line. Transitions are indicated in italics, while variants with MutPred scores above 0.5 are indicated in boldface.

The MutPred adjusted load hypothesis was previously described (Chapter 5: Venter et al., 2017), where MutPred scores of non-synonymous variants for a sequence are summed to obtain a total mutational load score. In the Venter et al. (2017) study, two adjustments were made to the total mutational load. Firstly, all variants with MutPred scores below the "actionable hypothesis" threshold (0.5) were excluded from the calculations, as these variants are unlikely to have a phenotypic effect, being common population variants likely to define major haplogroups. Mutational loads calculated with this adjustment are referred to as MutPred adjustment 1 loads in this thesis. Secondly, because of the rCRS position within the phylogeny, as well as the differences in the number of mtDNA variants between population groups, mutational load scores were also divided by the number of variants used to calculate loads, as a second adjustment. Both these steps supported minimisation of any bias that might be introduced by population stratification. However, because the cybrid cell lines used here were of similar haplogroups, this second adjustment step was not applied here. Table 7.2 gives a summarised look at the different genetic parameters important for this study.

To assess mitochondrial function, several parameters of mitochondrial respiration can be measured using the Seahorse XF^{e} 96 extracellular flux analyser, along with a Mito stress test. The process involves the measurement of the rate at which oxygen is consumed (OCR) by cells within each well of a Seahorse culture plate. After a basal respiration measurement is taken, a sequence of inhibitor injections first

inhibits ATP synthase to determine ATP production and proton leak, followed by the uncoupling of OXPHOS to measure maximal respiration rate. Finally, inhibitors of complex I and III are injected to completely block mitochondrial respiration; the remaining oxygen consumption is ascribed to non-mitochondrial respiration and is subtracted from all other measurements. Previous studies (Gómez-Durán et al., 2010) have shown that mtDNA copy number fluctuates greatly in cell cultures, but then stabilise after ~22 passages. Accordingly, the cybrid cell lines were cultured past 22 passages, while mtDNA copy number was monitored at certain time points. The Seahorse XF^e analyses were eventually carried out on cells after 35 passages. Unfortunately, even at this late state, RMCN still differed between passages within each cell line. Although RMCN is not expected to be exactly the same for all eight cybrid cell lines, there also were no consistencies in copy numbers in relation to each other. In other words, one cell line did not repeatedly have the highest copy number while another had the lowest. As an example, in Figure 7.2 it can be seen that whilst SA19 had one of the lowest RMCN at passage 26, it had by far the highest RMCN at passage 35, when mitochondrial respiration was assessed.

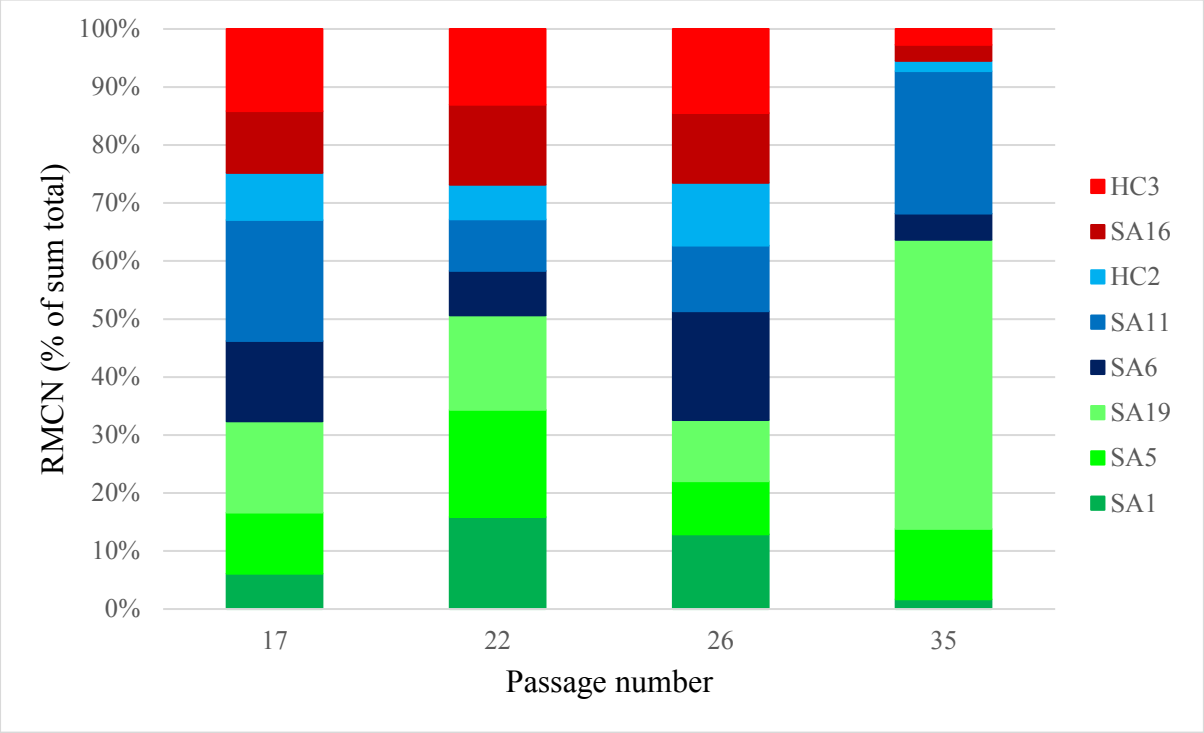


Figure 7.2: RMCN ratios between different cybrid cell lines. In this figure, the mtDNA copy number per cell is given as a percentage of the sum total of the entire group, to demonstrate the changes in ratio between different cybrid cell lines at various passages (time points).

Although an attempt was made to measure the RMCN at a consistent cell confluency (80-90%), at each time point, this was not always possible as not all cybrid cell lines had the same growth rate. mtDNA copy number has been shown to vary greatly at different stages of the cell cycle. Because non-confluent cell cultures are more likely to proliferate, it is possible that differences in confluency at the time that cells were harvested for RMCN assessment, might explain the differences in RMCN from one time point to the next. Despite this underlying concern, mitochondrial respiration of all eight cybrid cell lines was assessed at passage 35, using the Mito stress test.

Because the 143B osteosarcoma cell line is cancerous, it is more likely to rely on glycolysis for ATP production rather than OXPHOS (Crabtree effect). In an attempt to force cells to make use of OXPHOS, galactose-containing media was used instead of glucose-containing media in one assay plate when seeding cells 24 hours before the assay. The bio-energetic profile measurements of this plate were then compared to that of the other plate (cultured in glucose-rich media), seeded at the same time. Measurements of basal respiration, ATP production, proton leak and non-mitochondrial respiration in galactose-rich media, were all twice that of the same measurements on the same cell lines in glucose-rich media (results not shown). Interestingly, maximal respiration rates were only slightly higher in galactose-rich media than in glucose-rich media. Spare respiratory capacity is calculated as the difference in maximal respiration and basal respiration, and is an indication of the cell's ability to cope with stressful conditions. Because basal respiration rates were doubled in galactose-rich media but maximal respiration rates were not, the spare respiratory capacity of cells in galactose-rich media is comparably less. Table 7.3 gives the means and standard deviations of OCR measurements for each bio-energetic profile parameter, in each cybrid cell line, cultured in galactose-rich media. These measurements (from the assay plate with cells seeded in galactose-rich media), were used for the rest of this study. ATP production, maximal respiration, proton leak and non-mitochondrial respiration measurements all correlated significantly with basal respiration rates (Pearson's r = .98; .90; .52; and .67 respectively, all Ps < 0.0001). To simplify further statistical analyses, basal respiration rates were chosen as representative of mitochondrial function (OXPHOS) in these cells.

Table 7.2: Genetic parameters for each cybrid cell line

Cybrid cell line	Number of variants with MP scores <0.5	Number of variants with MP scores >0.5	Total number of variants with MP scores	MutPred total load	MutPred adjustment 1 load	RMCN
SA1	4	0	4	1.24	0	1260
SA5	5	0	5	1.56	0	9200
SA19	6	0	6	1.81	0	37768
SA6	4	1	5	1.86	0.63	3427
HC2	6	1	7	2.66	0.60	1351
SA11	5	1	6	2.19	0.66	18571
HC3	5	3	8	3.43	1.81	2058
SA16	4	4	8	3.96	2.59	2106

In this table, several genetic parameters relating to mtDNA are given. The number of non-synonymous variants with MutPred (MP) scores below or above the actionable hypothesis threshold (0.5) are given for each cybrid cell line used, as well as the total number of non-synonymous (MP-scoring) variants. MutPred adjustment 1 loads are mutational loads calculated by summing the MutPred scores of only variants with a score above 0.5. RMCN: relative mtDNA copy number refers to the number of mtDNA molecules per cell.

Table 7.3: Bio-energetic parameters for each cybrid cell line

Cybrid cell line	Basal respiration		ATP production		Proton leak		Maximal respiration		Spare respiratory capacity		Non-mitochondrial respiration		Coupling efficiency (%)	
	N	Mean ± SD	*N*	Mean ± SD	*N*	Mean ± SD	*N*	Mean ± SD	*N*	Mean ± SD	*N*	Mean ± SD	*N*	Mean ± SD
SA1	10	1.62 ± 0.21	9	1.07 ± 0.12	9	0.55 ± 0.12	10	1.42 ± 0.39	10	-0.20 ± 0.24	9	0.24 ± 0.09	9	66.31 ± 3.43
SA5	9	2.46 ± 0.11	9	1.98 ± 0.08	11	0.52 ± 0.12	11	2.69 ± 0.30	12	0.27 ± 0.30	11	0.52 ± 0.10	10	80.29 ± 2.72
SA19	12	1.09 ± 0.28	12	0.81 ± 0.18	11	0.35 ± 0.10	12	1.13 ± 0.35	12	0.04 ± 0.13	12	0.26 ± 0.36	10	68.76 ± 2.63
SA6	11	1.97 ± 0.30	12	1.66 ± 0.31	10	0.39 ± 0.06	12	2.17 ± 0.55	12	0.10 ± 0.47	12	0.44 ± 0.16	11	79.89 ± 2.43
HC2	9	2.84 ± 0.10	11	2.38 ± 0.16	10	0.47 ± 0.05	12	3.35 ± 0.71	11	0.65 ± 0.43	12	0.70 ± 0.08	12	81.48 ± 4.46
SA11	11	2.70 ± 0.23	12	2.31 ± 0.26	11	0.45 ± 0.05	11	3.49 ± 0.45	12	0.60 ± 0.48	12	0.71 ± 0.11	12	83.19 ± 1.63
HC3	10	1.57 ± 0.08	10	1.24 ± 0.06	10	0.33 ± 0.03	10	1.80 ± 0.22	10	0.23 ± 0.19	9	0.31 ± 0.04	8	78.70 ± 0.57
SA16	12	1.75 ± 0.32	12	1.37 ± 0.21	12	0.37 ± 0.13	11	1.90 ± 0.30	11	0.13 ± 0.10	12	0.40 ± 0.19	12	79.16 ± 3.53

In this table, the mean and standard deviation of OCR (pmol/min/Abs) of bio-energetic parameters as measured by the Seahorse XF^e analyser are given for each cybrid cell line used. The number of repeats, excluding outliers, used for each parameter is also given (*N).*

For this study, it was hypothesised that mildly deleterious mtDNA variants might be associated with changes in mitochondrial respiration; mildly deleterious variants being defined as those with MutPred scores of 0.5 or greater. As mentioned, MutPred adjustment 1 loads refer to mutational loads calculated by using only those variants with a MutPred score above 0.5, for each cybrid cell line. It was expected that basal respiration rates would correlate in some way with MutPred adjustment 1 loads, or differ between cybrid cell lines with zero, one or more than two high MutPred-scoring variants, if these variants were impacting on phenotype. Figure 7.3 is a graphical representation of the relationship between basal respiration and three genetic parameters: MutPred adjustment 1 loads; number of variants with MutPred scores *below* 0.5; and RMCN (for simplicity, the line in this graph illustrates log-transformed values of RMCN). Green bars represent basal respiration rates of cybrid cell lines that have no high MutPred-scoring variants; blue bars represent those rates of cybrid cell lines that each have one variant with MutPred scores > 0.5; and red bars represent basal respiration rates of cybrid cell lines carrying more than two variants with MutPred scores > 0.5.

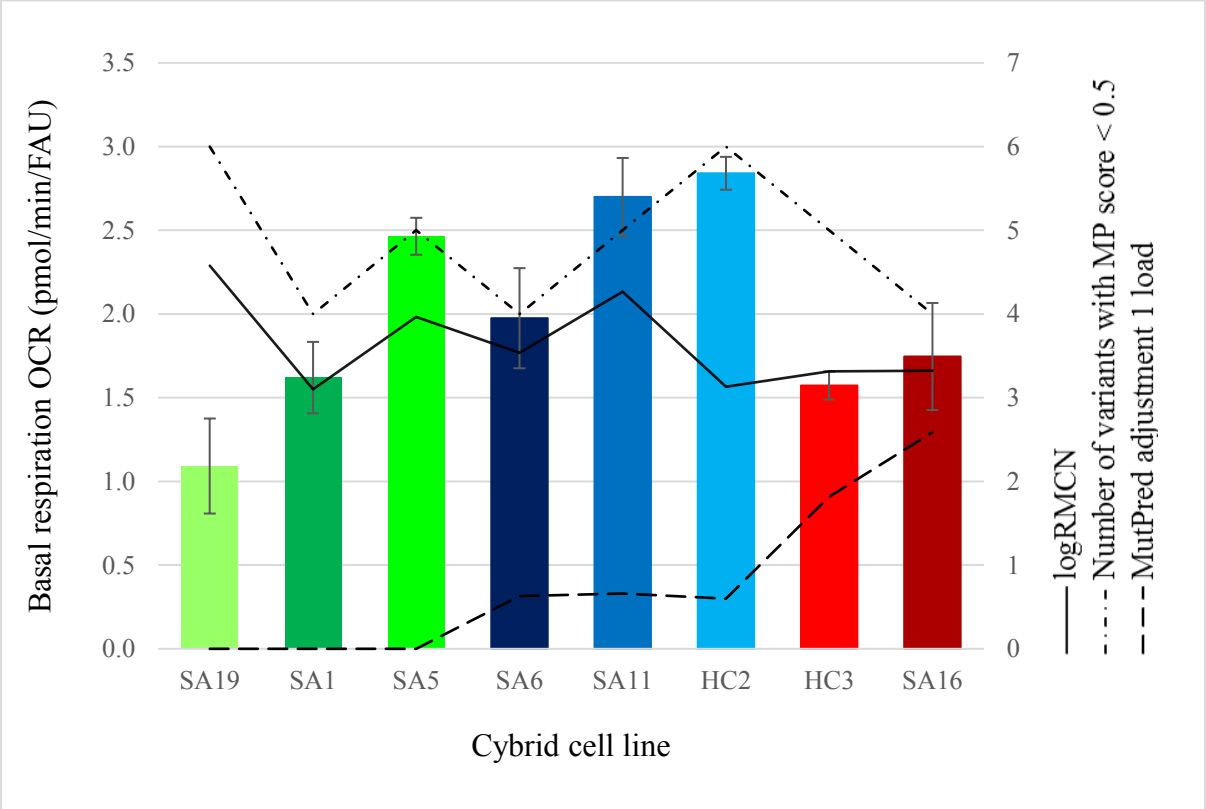


Figure 7.3: Relationship between mitochondrial basal respiration and genetic parameters. Means and SDs of basal respiration OCR measurements are indicated by columns, while genetic parameters are indicated by lines. The colour of the bar columns indicates the number of variants with MutPred (MP) scores above 0.5 in each cybrid cell line as follow: green = none; blue = one; red = more than 2. RMCN: relative mitochondrial DNA copy number.

At first visual inspection of the data, no clear relationship between MutPred adjustment 1 loads and basal respiration could be seen. While basal respiration rates of cybrid cell lines with the highest load scores (reds) were lower than in those with moderate load scores (blues), basal respiration rates of cybrid cell lines with load scores of 0 (greens) ranged from the lowest OCRs to relatively high OCRs. Aside from the number of high-scoring variants, the type of high-scoring variant did also not help to explain respiration level: very rare (Genbank frequency < 0.1%) high-scoring variants were present in SA6, SA16 and HC2 (Table 7.1), all of which had very different basal respiration rates. SA5 has one very rare, low-scoring variant. While both HC3 and SA16 had multiple high-scoring variants in complex I genes, SA11 and HC2 also had high-scoring variants in complex I genes. Having multiple high-scoring variants in complex V (SA16) also did not predict basal respiration rate very well, as three other cell lines lacking such combinations had lower respiration rates. No combination of non-synonymous mtDNA variants and mt-rRNA or mt-tRNA variants formed any clear relationship with basal respiration rates. As seen in Figure 7.3, two other genetic parameters appeared to more closely follow respiration trends: RMCN and number of variants with MutPred scores below 0.5. In fact, initial Pearson's correlation tests between all the genetic parameters listed in Table 7.2, and basal respiration rates, showed that only RMCN significantly correlated with respiration ($r = -.33$, $P = 0.002$). Curiously, this correlation was negative, while Figure 7.3 suggests a positive relationship. On closer inspection of the data, it became clear that the exceptionally high RMCN of SA19 (twice that of the next highest measurement) was responsible for this irregularity (Figure 7.4).

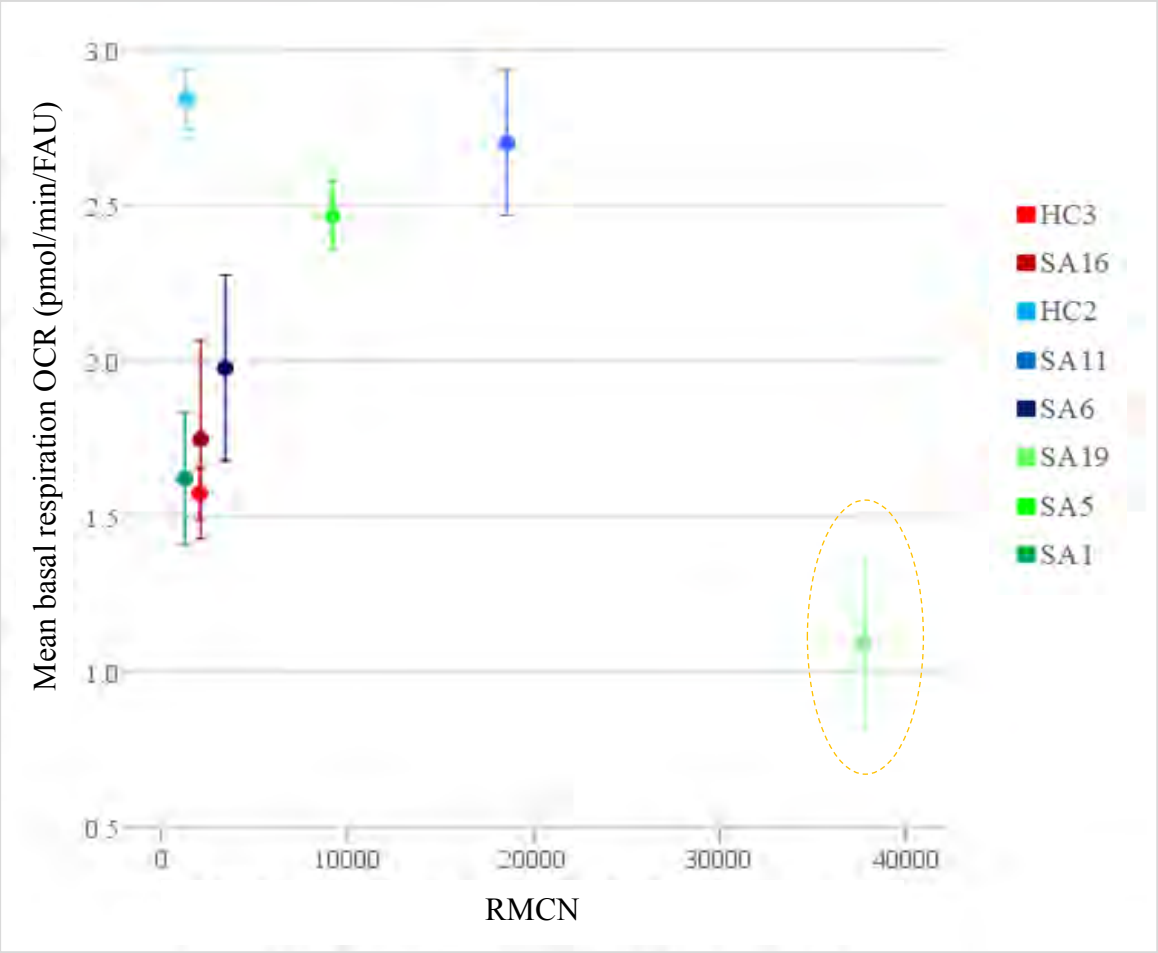


Figure 7.4: Graphical representation of the correlative relationship between basal respiration and RMCN. Mean basal respiration OCRs with SDs (error bars) are given in relationship to RMCN for each cybrid cell line. SA19 is marked with an ellipse.

Together with SA1, SA19 also had significantly lower coupling efficiency [F (7, 76) = 43.01, P < 0.0001] and significantly higher proton leak to basal respiration ratios [F (7, 72) = 51.1, P < 0.0001] than all other cybrid cell lines, indicating that OXPHOS coupling is somehow hampered in these two lines. Unlike SA19, SA1 however did not have significantly higher RMCN. The regulation of mtDNA copy number is not completely understood, but studies have shown that cellular proliferation could play a role (Trinei et al., 2006). mtDNA copy number is the highest in the late G_1 phase of the cell cycle, just before the transition into the S phase (when nDNA replication takes place before cell division). Since the S-phase requires an increased ATP output, it is thought that mtDNA is increased in the pre-S phase as preparation, before nDNA is replicated. Mitra et al. (2009) have also shown that during this stage in the cell cycle, mitochondria form large hyperfused networks. It is possible that the cells of SA19 were not as confluent as the other cell lines, and were proliferating at a higher rate. At the point of cell harvesting, many cells might have been in the G_1-to-S transitions phase, where mtDNA copy number is ramped up, but may not yet have translated this into increased ATP output. Another possibility is that despite an effort to keep all cell lines under identical conditions, including monitoring

of mycoplasma contamination, SA19 cells were somehow exposed to an unknown stressor. The significantly high RMCN, but with low mitochondrial respiration, appears anomalous and could not be disregarded. Indeed, in a similar previous study (van Dyk, 2016), which also included all the cybrid lines used in this study, SA19 was one of the best performing cell lines, with only SA5 and HC2 having higher basal respiration rates. For this reason, it was decided that SA19 should be excluded from further statistical analyses.

With SA19 now removed from the dataset, Pearson's correlation tests were again done to compare basal respiration rates with the genetic parameters listed in Table 7.2. From this, several genetic parameters correlated significantly with basal respiration rates. In Table 7.4 it can be seen that only the total number of non-synonymous (shown as variants with MP scores) variants and the MutPred total load did not significantly predict basal respiration rates. Note that the *P*-values given in this table are post Bonferroni correction for multiple testing.

Table 7.4: Pearson's correlations between mitochondrial basal respiration and genetic parameters

Genetic parameter	Pearson's coefficent (*r*)	Significance (*P*)	R^2
Number of variants with MP scores < 0.5	.67	< 0.00001	0.44
Number of variants with MP scores > 0.5	-.41	< 0.01	0.17
Total number of variants with MP scores	-.09	2.81	0.01
MutPred total load	-.24	0.25	0.06
MutPred adjustment 1 load	-.40	< 0.01	0.16
RMCN	.55	< 0.00001	0.30

In this table, Pearson's correlation coefficients (*r*), *P*-values and R^2 values are given where each genetic parameter was compared to basal respiration rates. *N* was equal to 72. A Bonferroni correction has already been applied to the *P*-values in this table. MP: MutPred, RMCN: relative mtDNA copy number.

From this re-analysed data, both the number of high MutPred-scoring variants, and the MutPred adjustment 1 load significantly correlated with basal respiration rates, in a negative direction. These two parameters could account for 17% and 16% of variation respectively, which is a noteworthy contribution. The negative relationships of these two genetic parameters with basal respiration rates

also make biological sense in a simplistic model, as these variants are assumed to be mildly deleterious. That is to say, an increase in the number of these variants or the mutational load from high-scoring variants lead to a decrease in basal respiration rates. It therefore seems, from this preliminary data, that the MutPred mutational load hypothesis could possibly be of biological significance.

RMCN significantly correlated with basal respiration rates (now in a positive direction), and accounted for 30% of variation, while the number of variants with MutPred scores below 0.5 could account for 44% of the variation. Both these parameters were significantly better at predicting basal respiration rates than the MutPred adjustment 1 load ($t_{Difference}$ = 10.2 for number of variants with MP-scores < 0.5; $t_{Difference}$ = 11.6 for RMCN; *P*s > 0.01). As mentioned, mtDNA copy number is likely regulated during the cell cycle, and higher copy numbers could translate into more OXPHOS complexes being assembled to achieve higher ATP turnover (Trinei et al., 2006). This would result in a positive correlation between RMCN and basal respiration as shown here; changes in mitochondrial respiration would then be a consequence of a change in mtDNA copy number. However, upregulated expression of mtDNA copy number could also be a consequence of aberrant mitochondrial respiration: Al-Kafaji and Golbahar (2013) have shown that oxidative stress induced by high glucose concentrations increased mtDNA copy number in human mesangial cultured cells, possibly to compensate for mtDNA oxidative damage. Although cell growth was limited in galactose-rich media when compared to the glucose-rich assay plate, it is not known whether reactive oxygen species or mtDNA were increased in the cybrid cells used here, as this was not measured. Indeed, Rossignol et al. (2004) have shown that mtDNA copy number did not significantly differ between HeLa cells grown in galactose- or glucose containing media, which might also be the case for the 143B cells used as base cell line in this study. Turning to the number of variants with MutPred scores below 0.5, as non-synonymous variants with low MutPred scores are more likely to be common population variants, the significant correlation found here would rather support the hypothesis that, in terms of mtDNA variants, mitochondrial respiration is more influenced by haplogroup background than rarer, mildly deleterious variants. Although this result was not expected, it is not in contradiction to existing literature.

In a series of papers, Kenney et al. (2014a; 2014b) showed that: i) while ATP production and spare respiratory capacity were comparable between haplogroup H cybrids and haplogroup J cybrids, the OCR to ECAR ratio (OXPHOS:glycolysis) was higher in haplogroup H cybrids (Kenney et al., 2014b); and ii) haplogroup L cybrids had lower mtDNA copy number and decreased ATP and ROS production, but higher expression levels of nine mtDNA-encoded OXPHOS complexes, when compared to haplogroup H cybrids (Kenney et al., 2014a). However, is interesting to note that when MutPred adjustment 1 loads are calculated for the cybrid cell lines used in the above studies, haplogroup J cybrid cell lines have load scores of 3.1, 1.2 and 0.6, while all haplogroup H cybrid cell lines used by Kenney et al. have no variants with MutPred scores above 0.5 (i.e. MutPred adjustment 1 loads of 0).

Haplogroup L cybrids also had several variants with MutPred scores above 0.5, with MutPred adjustment 1 loads of 0.6, 1.7 and 2.9 for the three cell lines used. Thus, any comparisons between haplogroup J or L cybrid cell lines and haplogroup H cybrid cell lines would also be a comparison between high and low mutational load scores. In the above studies, that would result in cybrids cell lines with high mutational load scores (haplogroups J and L) having lower bio-energetic profile parameters than those with scores of 0 (haplogroup H). Nevertheless, this was not considered in the Kenney et al. studies and these authors concluded that the observed differences in their study (which include the altered expression of eight nuclear encoded genes, in addition to the changes mentioned above) are "not due to rare mtDNA variants but rather the combination of SNPs representing the J versus H haplogroups" (Kenney et al., 2014b). The limited cybrid data presented here supports this observation to some extent.

Next, the cybrid cell lines were grouped into two of six categories: for variants with MutPred scores below 0.5, cell lines carrying either 4, 5 or 6 variants each; for variants with MutPred scores above 0.5, cell lines carrying either 0, 1 or more than 2 variants each (Figure 7.5). The differences in basal respiration rates were tested using one-way ANOVAS (Table 7.5). For the number of variants with MutPred scores below 0.5 categories, all three groups differed significantly from each other. However, from Figure 7.5A it can be seen that basal respiration rates of HC3, having 5 such variants, is more comparable with basal respiration rates of cybrid cell lines having only 4 variants. Still, the significant differences between these groups, would suggest that having more variants with MutPred scores below 0.5 results in higher mitochondrial respiration (Figure 7.5A). However, these variants are not believed to have any significant impact on structure or function. The way in which these variants could play a role in mitochondrial respiration rates would have to be more indirect. Cross-talk between mitochondrial DNA and nuclear genes has been suggested as a possible way for these population variants to impact on mitochondrial function (Picard et al., 2016). This result is also not consistent with the work of Kenney et al., as haplogroups J and L have several more low-scoring, non-synonymous variants when compared to haplogroup H, but were shown to have lower measurements of mitochondrial function than haplogroup H (i.e. a negative relationship between the number of low-scoring variants and mitochondrial function measurements, as opposed to the positive relationship shown in the current study).

When considering the number of mtDNA variants with a MutPred score below 0.5 (which are assumed to be common population variants) as a possible predictor of mitochondrial function, one should be careful to keep bias introduced by population stratification in mind. Differences in the number of such variants could simply be an artefact of the position of the cell line mtDNA sequence in the phylogeny, in relation to that of the rCRS. A way to test whether this specific type of bias has had any influence on the data in question, would be to call variants against the reconstructed sapiens reference sequence

(RSRS), which roots the phylogeny in such a way that it shows no bias in phylogenetic distance to any modern haplogroups (Behar et al., 2012). When this check was applied for the eight cybrid cell lines used in this study, three variants (m.8860A>G, m.14766C>T and m.15326A>G) previously called against the rCRS were excluded from variant lists. However, given the closely related haplogroups of the cybrid cell lines in this study, all three variants were previously included in all eight cell lines, and as such, the exclusion of the variants was uniform across all eight cell lines. Consequently, the statistical relationships between the number of variants with MutPred scores below 0.5 and basal respiration rates were unaffected.

For the number of variants with MutPred scores > 0.5 categories, cybrid cell lines having one such variant significantly differed from both other groups. Yet, in Figure 5B it can be seen that basal respiration rates of SA6 were not consistent with those of other cybrid cell lines in the same group (having one high-scoring variant). The single high-scoring variant present in SA6 is very rare, and not present on Phylotree or GenBank. It could be possible that this variant is deleterious enough to be as biologically significant as having more than one high-scoring variant. Groups with cybrid cell lines that have none or more than two variants with a MutPred score above 0.5 did not significantly differ from each other, since SA1 had very low basal respiration rates. As mentioned previously, SA1 also had significantly lower coupling efficiency compared to other cell lines (66% in SA1 vs ~80% in other cell lines). Whether this lowered coupling efficiency is a direct consequence of mtDNA variants (other than high-scoring variants, as SA1 had none), or rather some other biological or methodological factors, is unknown. If the basal respiration rates for SA5 are taken to be more representative of cell lines carrying no high MutPred-scoring variants, then the data presented in figure 7.5B would suggest that mitochondria can tolerate small aberrations in OXPHOS complexes, introduced by single high MutPred-scoring variants, but that some threshold is exceeded in cell lines carrying several such variants (cell lines HC3 and SA16), disrupting mitochondrial respiration. However, whether these decreased respiration rates are detrimental to cell viability, and by extension, health, is not definite. Subtle changes in coupling efficiency have been implicated in survival rates after sepsis (decreased coupling efficiency might be beneficial by aiding in increased heat production during sepsis), in a prospective study comparing haplogroup H patients to patients from other haplogroups (Baudouin et al., 2005). It has also been suggested that having “leaky” mitochondria could protect against diseases such as diabetes type 2 in the presence of an over-supply of nutrients: sporadic alleviation of the membrane potential over the inner membrane via proton leak, ensures that the OXPHOS system does not reach a static head equilibrium, which would increase electron leak, resulting in increased H_2O_2 production and consequent downstream damage by oxidative stress (Fisher-Wellman & Neufer, 2012). Including measurements for cell viability and oxidative stress markers in future cybrid studies could help to further elucidate this putative relationship between mtDNA variants, mitochondrial respiration and health.

Table 7.5: Comparisons of mean mitochondrial basal respiration between different mtDNA variant groups

	Mitochondrial basal respiration OCR (pmol/min/Abs)						
	N	Mean ± SD	*F* (df_M, df_R)		*N*	Mean ± SD	*F* (df_M, df_R)
Number of variants with MutPred scores < 0.5				*Number of variants with MutPred scores > 0.5*			
4	33	1.78 ± 0.31	27.7 (2, 69)*	0	19	2.02 ± 0.46	26.9 (2, 69)*
5	30	2.25 ± 0.52		1	31	2.48 ± 0.45	
6	9	2.84 ± 0.10		> 2	22	1.67 ± 0.25	

In this table, mean basal respiration rates of different mtDNA groups were compared using one-way ANOVAS. Means and standard deviations (SD) are given with *F*-ratios (*F*) and degrees of freedom for the effect of the model (df_M) and residuals (df_R). Asterisks indicates differences significant at the < 0.0001 level.

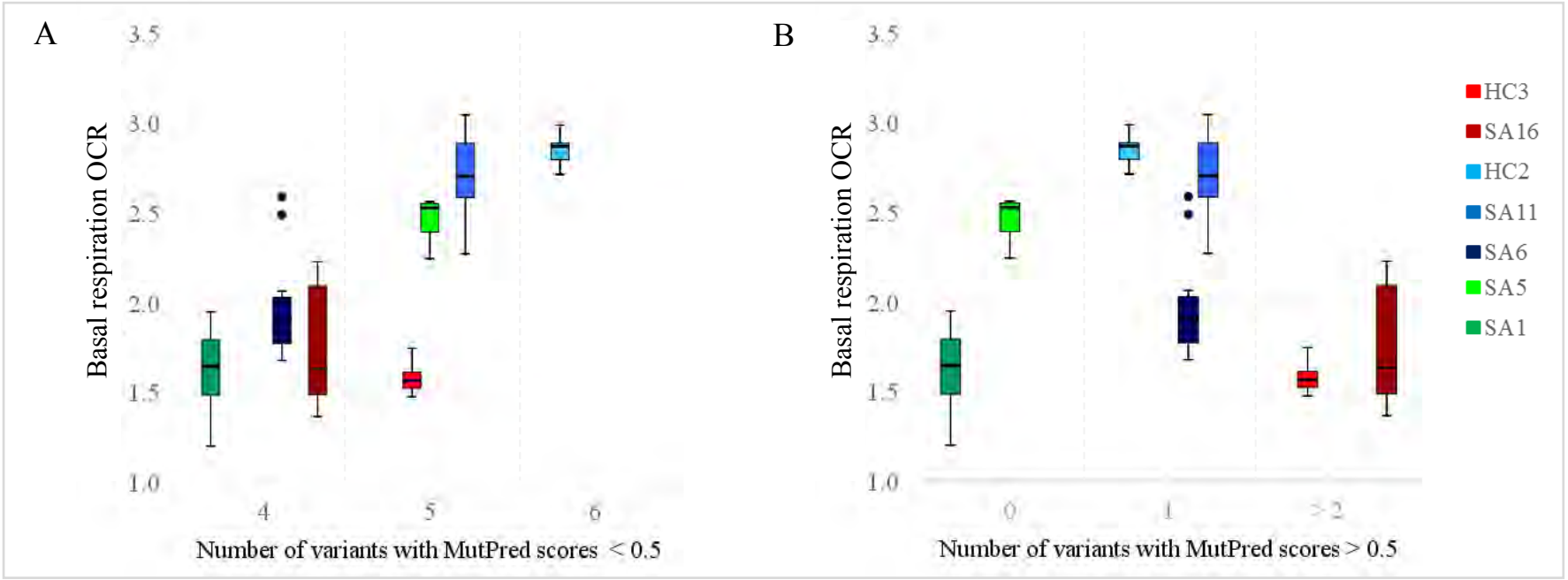


Figure 7.5: Comparison of basal respiration between cybrid cell lines within different mtDNA variant groups. Box-plots indicated the means, 25% and 75% quantiles as well as standard deviations of basal respiration for each cell line, grouped according to the number of variants with MutPred scores below 0.5 (A) or above 0.5 (B), that each cybrid cell line carried.

7.4 CONCLUDING REMARKS

This cybrid cell line study aimed to test whether those with MutPred scores above 0.5 (mildly deleterious variants), either in isolation or in combination, impact on mitochondrial respiration. Although the number of variants with MutPred scores above 0.5, as well as MutPred adjustment 1 loads significantly correlated with mitochondrial respiration, correlations between RMCN and the number of variants with MutPred scores below 0.5, with mitochondrial respiration was more pronounced. While some of the results here are consistent with previous studies (low MutPred-scoring/population variants have a higher impact on mitochondrial function than high MutPred-scoring/rare variants), others are not (in this study a positive relationship between the number of low MutPred-scoring variants and mitochondrial respiration was found, while the Kenney et al. studies suggest a negative relationship). Although the data presented here reveal some interesting insights into mtDNA variation and mitochondrial function, there are some important shortcomings: i) the statistical power of this study was limited by methodology requirements, such as the number of repeats needed for a cell line, and the fact that all cell lines needed to be analysed in the same experimental run; ii) more refined selection criteria, such as the inclusion of cybrids cell lines containing multiple very rare mtDNA, non-population variants or exactly matched haplogroups, was restricted by sample availability; iii) heteroplasmy data was not available for all samples used, and therefore the possible influence of such variants could not be investigated; iv) despite the cells being cultured past an advanced number of passages, the RMCN was still not stable, and could have affected mitochondrial respiration significantly; v) factors such as mtDNA damage, ROS production, cell viability and mtDNA gene expression were not measured and could not contribute to further interpretation of the data. Thus, further investigations are warranted, in studies where these shortcomings are addressed. It would also be of great advantage to the validity of such studies if their results can be duplicated with several other haplogroups.

Chapter 8: Summary and conclusions

In this chapter the rationale, theoretical and experimental approaches, as well as the key findings of the study, are summarised, critically evaluated and assimilated to provide a basis for the conclusions that were finally made.

8.1 BIOLOGICAL RATIONALE FOR CONSIDERING MTDNA VARIATION IN HUMAN DISEASE

The endosymbiotic theory states that the organelles mitochondria and chloroplasts are derived from ancient endosymbionts of prokaryotic organisms (Mereschkowsky, 1910). The attainment of an energetically favourable aerobic electron transport system would have given the host organism a remarkable evolutionary advantage, making possible the complexity of contemporary life. Indeed, mitochondria in most eukaryotic cells produce ~90% of cellular energy (Chance et al., 1979), but are also intricately involved in essential cellular processes by: i) influencing homeostasis and adaptive stress signalling pathways (including cell proliferation, differentiation and apoptosis); ii) housing several biosynthetic pathways (including for pyrimidines, lipids, heme and Fe-S clusters, and neurotransmitter precursors); iii) regulating levels of a wide range of cellular compounds (metabolites, amino acids and enzyme co-factors); and iv) modulating Ca^{2+} flux (reviewed by Nunnari & Suomalainen, 2012). The production of both ATP and ROS by the OXPHOS complexes is central to most of these functions. This broad extent of mitochondrial influence has, unsurprisingly, lead to the implication of mitochondrial *dys*function in not only rare mitochondrial disease, but also in metabolic and degenerative common complex disease (Nunnari & Suomalainen, 2012; Wallace, 2013). During the transition from endosymbiont to organelle, the precursor of the mitochondrion transferred most of its genes to the host organism, retaining only 38 mitochondrially encoded genes in humans. The loss of mitochondrial genes to the nucleus has several, widely accepted advantages, which are typically centred around the preservation of genetic integrity (Johnston & Williams, 2016). The factors that determine which genes are retained within the mitochondrion are the subject of an ongoing debate (Allen, 2015). A very recent study by Johnston and Williams (2016) provided robust statistical support for three factors to greatly influence the retention of mitochondrial protein coding genes in eukaryotes: protein hydrophobicity (it might be difficult to correctly transport highly hydrophobic proteins into mitochondria from elsewhere in the cell), gene GC-content [high GC content provides thermodynamic stability to the mtDNA molecule (Samuels, 2005), which could serve as a selective force], and energetic centrality (retained genes make up the core subunits of the ETC). The benefit of energetic centrality is described by the CoRR (co-location of gene and gene product for redox regulation of gene expression) hypothesis (Allen, 1992). The CoRR hypothesis states that the redox state of locally encoded gene

products (such as ETC subunits in the mitochondrion) can serve as a regulatory control for the expression of those genes (Allen, 2015). The advantage of such a system would be that the composition of enzyme complexes is directly regulated by the redox environment within the mitochondrion, allowing for acute and fine bioenergetic modulation (Allen, 2015; Johnston & Williams, 2016). In another study, Maier et al. (2013) illustrated that the two rRNA subunits encoded by mtDNA are also central to the assembly of mitochondrial ribosomes. The 22 tRNAs encoded by mtDNA is the minimum subset necessary to read all the mtDNA codons (Florentz et al., 2003). From all of this, it should be clear that: i) the genes encoded by mtDNA are those most essential to optimal mitochondrial function and regulation, and ii) mitochondrial function in turn is essential to optimal cellular function, and by extension, human health. This lends a solid biological rationale for also considering variation in mtDNA (despite its relatively small size when compared to nuclear DNA) when investigating genetic factors involved in a wide range of rare and common diseases, including cardiometabolic and other late-onset diseases. Indeed, as discussed in detail throughout this thesis, many studies have done exactly that in the last decade or so, utilising several approaches (see Chapter 2). However, when investigating mtDNA variation, irreproducible, inconsistent and contradicting reports are common in literature (Salas & Elson, 2015), indicating the need for alternative approaches when considering mtDNA variation in association with disease. It was therefore also set as a specific goal of this study to further develop and implement such a novel approach, the core concept of which was proposed by Elson et al. (2006), and which is presented here in the form of the "MutPred adjusted load" hypothesis.

8.2 PROBLEM STATEMENT AND AIM

With the growing epidemic of cardiometabolic disease, especially in rapidly urbanising African populations, it becomes essential to identify all factors that could be involved in the onset and progression thereof (Mensah et al., 2015). Identification of genetic risk factors, which then also include mtDNA variation, could help in understanding the differences between population groups in disease susceptibility and outcome, in addition to the well-documented environmental and life-style factors known to play a significant role (Malan & Malan, 2016). Sub-Saharan African populations are under-represented in genetic databases, as well as genetic studies concerning cardiovascular disease. Even in well-characterised studies on mtDNA variation in disease done mostly on Caucasians, current approaches have often delivered results that often include inconclusive and contradicting reports (Salas & Elson, 2015). Thus, the aim of this study was to determine whether mtDNA variants play a significant role in the presentation or severity of cardiometabolic disease in a bi-ethnic cohort, using current and novel investigative approaches.

8.3 OBJECTIVES: RESULTS AND LIMITATIONS OF THIS STUDY

8.3.1 *First objective: generating mtDNA sequences for African populations*

The first objective of this study was to "generate and contribute complete mtDNA sequences for an under-represented population group (Sub-Saharan Africans), as well as a Caucasian group from the same geographical area".

Of the 409 participants in the SABPA cohort, all from the same geographical area (North-west province, South Africa), 200 were Black South Africans and 209 were Caucasian South Africans. Technical aspects that each contributed to a few cases of sample fall-out were difficulty in obtaining good quality isolated DNA samples, failure to obtain amplicons of both PCR fragments, or loss of genetic material during clean-up steps. In the end, complete mitochondrial sequences were produced and reported for 194 Black South Africans and 169 Caucasian South Africans. Currently on GenBank, haplogroup L3 is the best-represented African haplogroup. The majority of African sequences generated in this study were from haplogroup L0 (61.5%), followed by L2 (17%), L3 (16.5%), L1 (4.6%) and lastly a single L4 sequence. Non-African sequences generated in this study were predominantly from haplogroups H (39%), U (21%), T (14.3%) and J (%), with single-digit representations for haplogroups B, I, K, M, N, R and W. Variant lists as called against the rCRS are published as supplementary material in Venter et al (2017), and complete sequences will shortly be made available to both GenBank and Phylotree. The first objective of this study was thus successfully achieved.

8.3.2 *Second objective: Disease-associated mtDNA variants in hypertension*

The second objective of this study was to "test whether disease-associated variants or known pathogenic mutations are more frequently found in hypertensive groups than control groups".

This objective was addressed in Chapter 4, in particular Section 4.3.1, by assessing the frequency of disease-associated variants, as reported by MITOMAP, in hypertensive groups when compared to normotensive groups. As an unbiased approach, *all* disease-associated variants reported by MITOMAP as of August 2016, were used for frequency analyses, without evaluating the evidence of the reported associations. However, it is well known from literature that many of these variants have weak evidence of disease association, and not surprisingly, most of the variants listed here had an even distribution among phenotype groups. From the full list, four disease-associated mtDNA variants were identified (m.961T>C, m.5460G>A, m.5911C>T and m.10398A>G) based on a distribution tendency towards segregating with disease. However, none of these distributions were statistically significant after a Bonferroni correction for multiple testing, even though two of these variants (m.961T>C and

m.5911C>T) were only found in haplogroup L hypertensives. Taken together, this highlights some of the difficulties faced by current approaches such as haplogroup association studies, and the need for alternative approaches when mtDNA is investigated in relation to disease. Having to do a number of Fisher's exact or chi square tests to evaluate each individual variant imposes a large statistical burden on studies, as corrections for multiple testing need to be done. Frequency analyses are also crude in the sense that they are essentially quantitative in nature, and the predicted, or known biological structural and functional impact of a variant is not necessarily considered. In contrast, the MutPred mutational load hypothesis reduces this statistical burden by combining individual variants into a single continuous metric (mutational load), and is also qualitative in nature, as it incorporates the pathogenicity scores of variants involved. However, being aware of the limitations imposed by the moderate cohort size, as well as current approaches, the objective here was specifically not to do a haplogroup association study, or attempt to investigate individual variants in relation to hypertension. Rather, the set objective was to simply assess whether disease-associated variants as a group were found more frequently in hypertensives than in normotensives, as a way of exploring the data. It was concluded that this was not the case. Therefore, the objective was successfully achieved.

Interestingly though, three variants that are considered to be known pathogenic mutations were also identified in this cohort. The first, m.11778G>A, is a well-known LHON mutation (Yu-Wai-Man & Chinnery, 2000 [Updated 2016]), but was found at only 16% heteroplasmy level in one participant and is unlikely to be physiologically relevant in this case. Two mt-tRNA mutations, m.5814T>C and m.14687A>G, were found in four participants. These mutations were classified as pathogenic using the pathogenicity scoring system described in Yarham et al. (2011). However, as discussed in Chapter 4, the evidence of pathogenicity for both these variants now seems questionable, and it is likely that both these variants, given their GenBank frequencies (between 0.1% and 1%) and presence on Phylotree, should be reclassified as rare population variants. The re-evaluation of these two variants does raise an important issue concerning the classification of disease-causing variants: the functional evidence for pathogenicity was reported more than a decade ago. It is thus probable that, at the time, both these variants were thought to be very rare, as would be expected for a mutation causing primary mitochondrial disease (mitochondrial encephalopathy and mitochondrial myopathy in this case). Similar observations have been reported in an investigation of mtDNA in African patients with mitochondrial disease (van der Walt et al., 2012). However, as more sequences have been added to databases and the human mtDNA phylogeny has been refined, these variants have been found in hundreds of assumedly non-diseased study participants. Because genetic counselling can have an enormous impact on a patient's life, such as choosing not to have a child, or choosing mitochondrial replacement therapy for *in vitro* fertilisation, it is important to re-evaluate the pathogenicity status of mutations on a regular basis. Indeed, a preliminary scan of mt-tRNA mutations previously classified as pathogenic, using the Yarham scoring system, identified two other mutations, m.5628T>C and

m.7472T>C, as potential candidates for re-evaluation, based on GenBank frequency and their presence on Phylotree as haplogroup markers. It would be a useful exercise to systematically inspect the evidence for pathogenicity put forth for these and other mutations that are routinely used during genetic counselling.

8.3.3 *Third objective: mitochondrial tRNA and rRNA variants in hypertension*

The third objective of this study was to "test whether mitochondrial tRNA and rRNA variants are more abundant in hypertensive groups compared to control groups".

This objective was addressed in Chapter 4, in particular Section 4.3.2. It was argued that mt-tRNA variants are good candidates for being linked to disease, as they occur more at the tips of the phylogeny than within the deeper branches, indicating an ongoing selection against these variants (Yarham et al., 2010). Also, compared to mutations in mtDNA protein encoding genes, pathogenic mutations of both mt-tRNA and mt-rRNA genes have a much wider impact on mitochondrial proteins, as the products thereof are involved in translation and transcription of all mitochondrial protein genes (Elson et al., 2015; Florentz et al., 2003). However, again taking into consideration the limitations imposed by the moderate cohort size, the objective here was not to attempt to identify specific known or novel mt-tRNA or mt-rRNA variants in association with hypertension, but simply to test whether (consensus) mt-tRNA or mt-rRNA variants that were rare on either GenBank (< 0.1% frequency) or rare within each gender/background group (only occurring once per group) were more abundant as a group, within hypertensives compared to normotensives. As discussed in Chapter 4, when taking into consideration the size of each group, the differences in frequency of these variants across hypertensive and normotensive groups were not statistically significance. Considering these findings, this objective was successfully achieved.

The fact that rare mt-tRNA and mt-rRNA variants are more abundant in haplogroup L participants, might be a result of the under-representation of African sequences in databases. Variants might be rare in GenBank, but well represented within unexplored African populations. This can make it difficult when novel disease-associated variants are sought, and highlight the need for more sequencing to be done in SSA populations.

8.3.4 *Fourth objective: Non-synonymous mtDNA variants in hypertension and hyperglycaemia*

The fourth objective of this study was "Using a new approach (a modified mutational load hypothesis), investigate the possible role of non-synonymous protein coding mtDNA variants *in the presentation of hypertension or hyperglycaemia"*.

This objective was addressed in Chapter 5, which is presented in this thesis as a published, peer reviewed paper (Venter et al., 2017). Here, only consensus and non-synonymous (protein coding) mtDNA variants were considered. First the data was explored by comparing the number of very rare (GenBank frequency of < 0.1%) and common (GenBank frequency of > 1%) non-synonymous population alleles unique to each group as defined by haplogroup background (L or MN), gender and phenotype. Although hypertensive haplogroup L participants had more very rare variants than haplogroup MN participants in any groups, no statistically significant differences were found. As mentioned, it is more likely that these high rare allele counts in hypertensives are a result of the under-representation of African sequences on GenBank, as well as the clinical cut-off points for hypertension classification that are possibly unfit for use in a uniform manner across different population and gender groups.

Next, MutPred pathogenicity scores were assigned to all non-synonymous mtDNA variants. To minimise the influence of population variants, which are proposed to be of little biological relevance and are likely to have low MutPred scores, variants with scores below the "actionable hypothesis" threshold (0.5) (Li et al., 2009) were excluded from further analyses. MutPred total (mutational) loads were calculation by summing the MutPred scores for the remaining variants for each participant. However, since the number of mtDNA variants a sequence can have is also an artefact of its position in the phylogeny in relation to that of the rCRS, and since some population variants do have MutPred scores above 0.6, MutPred adjusted loads were calculated by dividing MutPred total loads by the number of variants used, as a second adjustment. These corrections were made even though comparisons were not made between haplogroup L and haplogroup MN groups, since the genetic variation within African haplogroups alone can be quite large, and a homogenous group of "haplogroup L" participants is not guaranteed. Alternatively, this issue could be mitigated by the use of the RSRS instead of the rCRS. This would, however, bring about difficulties in relating data to existing literature, and so the consensus in the field is the continued use of the rCRS (Bandelt et al., 2013). Comparisons of MutPred adjusted loads between hypertensive and normotensive, or healthy, pre-diabetic and diabetic participants within gender/mtDNA background groups did not reveal any statistically significant results. Participants in disease groups (hypertensive, pre-diabetic or diabetic) were also no more likely to have variants with MutPred scores above 0.6 than control groups.

The MutPred adjusted load hypothesis was proposed as an alternative to the troubled haplogroup association approach. As mentioned, the combination of several characterised (pathogenicity scored) individual variants into a single qualitative metric allows for the use of simple parametric statistics and reduces the need for multiple testing corrections, which deliver more statistical power. Also, the two adjustment steps that were applied during the calculation of the final mutational loads used aim to limit the possible bias introduced by population stratification in analyses. From a *genetics* point of view, this means that it would have been acceptable to perform the various statistical analyses described in this thesis, where MutPred adjusted loads were used, across the different population groups. However, this was not done and the cohort was divided into four groups based on gender and mtDNA background (haplogroup), because of population stratification from a *phenotypical* point of view: great disparities exist in blood pressure and blood glucose measurements between both male vs female groups, and African vs Caucasian groups. In the SABPA study, many other biochemical measurements (Malan et al., 2015; 2017), as well as metabolic profiles (van Deventer, 2015) differed significantly between the four gender/mtDNA background groups. As mentioned several times in this thesis, the high percentage of haplogroup L participants that are classified as hypertensive, in a cohort that did not seek to specifically recruit diseased participants, in the author's opinion, leads to uncertainty on the validity of the universal use of current clinical threshold values to classify hypertension. Possibly, these values should be tailored to be gender and population specific. The lack of large and comprehensive CVD studies in African populations does not make such an adjustment to clinical threshold values easy, and should be addressed so that population specific, outcome-based reference values can be established. As a similar issue, Hoebel et al. (2013) proposed the use of different waist-circumference cut-offs associated with metabolic syndrome for use in African and European males and females. The 2013 ESH/ESC guide lines for the management of arterial hypertension does not list different reference values for males and females, despite the tendency for males to have higher blood pressure than females in pre-menopausal states (Piepoli et al., 2016). Thus, an alternative cut-off point, based on findings reported by Hermida et al. (2013) was used in this study to classify females as hypertensive or normotensive. Of course, even "correct" clinical cut-off points can be seen as arbitrary thresholds when used in research studies such as this one, as participants with measurements close to these thresholds might differ from each other with only a few points (mmHg for ABPM), but be classified into different phenotype groups. Clearly, such cases, which are likely to be much more abundant than cases on the extreme ends of the spectrum, would influence group averages and might mask actual differences between participants with very high or very low measurements. For this reason, correlation analyses are more likely to reflect the true relationships between genetic parameters and these phenotypical measurements (and were also used in this study).

Despite the fact that the cohort had to be divided into four groups and analysed separately, the study was comparatively well-powered. When the proposed hypothesis is considered where only variants with MutPred scores above 0.5 are included, and the position on the phylogeny is taken into account (see Table 4 in Chapter 5), the study had 80% power at the 0.05 significance level to detect a 0.106 difference in MutPred adjusted mutational loads between those with high blood pressure and those with blood pressure on the normal range, with 95 participants in each group. A 0.106 difference equates to approximately a 24% difference between the cases and controls. While this is a large difference, this still compares favourably to the traditional haplogroup association method as described in Samuels et al., (2006):

"*Studies with only 500 cases and 500 controls would have only 90% power to detect >35% increase in the frequency of haplogroup H, corresponding to an OR >1.75. For less common haplogroups or more-subtle changes in haplogroup frequency, even greater sample sizes will be required*"

As such, the methods proposed here, achieved a similar power to the traditional haplogroup methodology with approximately one fifth of the sample size. The use of a single, continuous genetic variable (MutPred adjusted load) also made it possible to test whether variation in mutational load values was associated with an increase in symptom severity, using correlation analysis. If larger sample sizes were to be used, even smaller differences between cases and controls could be detected. This attests to the usefulness of the concept of the MutPred mutational load hypothesis as an investigative tool when mtDNA variation in disease is considered. Thus, despite the limitations imposed on this study by the discrepancies in phenotype presentation between different gender and mtDNA background groups, this objective was successfully achieved.

8.3.5 *Fifth objective: Non-synonymous mtDNA variants in oxidative stress and inflammation*

The fifth objective of this study was: "using a new approach (a modified mutational load hypothesis), investigate the possible role of non-synonymous protein coding mtDNA variants in *altered levels of oxidative stress and inflammation indicators"*.

This objective was addressed in Chapter 6 using the same approach as described in detail in Chapter 5 and briefly in the previous section here, and presented in article format. From the previously described analyses, it was concluded that non-synonymous variants with MutPred scores above 0.5 did not play a significant role in hypertension or hyperglycaemia. Because blood pressure and blood glucose levels are both strictly controlled within humans, the question remained whether the previous result was

because other, more influential regulatory systems such as the sympathetic nervous system, could overcome or out-weigh comparably smaller changes resulting from downstream processes of mtDNA variation. In Chapter 2, inflammation, to which mtDNA contributes both via ROS and independently of ROS, was considered as a unifying factor involved in the aetiology of many CVD conditions. As such, measurements for seven indicators of oxidative stress and inflammation were selected from the data available on the SABPA cohort. All these measurements could putatively be influenced in a direct way by mitochondrial dysfunction, and so the role of non-synonymous mtDNA variants in these measurements were assessed using the MutPred adjusted load hypothesis. As with the previous statistical analyses done for blood pressure and blood glucose measurements, the measurements for markers used in this part of the study (ROS, NO, 8-OHdG, TBARS, GSH, CRP, and TNFα) also differed significantly between gender/mtDNA background groups. Statistical analyses were done separately for each group, and no statistically significant relationships were found. There were also no significant differences in the mean measurements between groups who had one or more mtDNA variants with a MutPred score above 0.5, and those who did not. Limitations of this part of the study could be those imposed by the accuracy of methods used to measure the biological parameters: for example, both ROS and NO bioavailability are measured indirectly, while TBARS was measured in urine and not blood. All seven indicators were also adjusted using age, BSA, TEE, serum cotinine (as indication of tobacco use) and cGGT (as indication of alcohol use), and in some cases CRP. Objectively measured lifestyle factors were obtained for the SABPA cohort and is a strength in the study (Hamer et al., 2011), but still, as is often the case for assays in biological samples, all these measurements have limitations in sensitivity and specificity, as well as inherent inter and intra-assay variability, which can be exacerbated by applying statistical adjustments using multiple factors. Another limitation could be that the cohort itself is not a severely diseased one, and it is possible that no significant results were found because disease progression is not advanced enough in the SABPA cohort to demonstrate the proposed impact of high MutPred-scoring variants. Also, it could be possible that such an impact exists, but is very small, and that the statistical power of this study was not sufficient to detect it, or that impact of environmental influences greatly out-weight mtDNA genetic risk factors in relation CVD susceptibility. Sympathetic hyperactivity may influence disease progression, and indeed was related to cardiac injury and blood pressure increases in Black South Africans SABPA participants (Malan et al., 2017). Since no direct measurement of mitochondrial function could be done for this cohort, is was not possible to know whether mitochondrial dysfunction was present in any of the participants. However, from the results presented in Chapter 6, it seems unlikely for this cohort, that non-synonymous mtDNA variants play a role in the selected oxidative stress and inflammation markers as measured for this study. As such the fifth objective of this study was successfully reached.

8.3.6 *Sixth objective: mtDNA variants in mitochondrial respiration*

The sixth objective of this study was to "investigate the influence of selected mtDNA variants on mitochondrial respiration in cytoplasmic hybrid cells".

This objective was addressed in Chapter 7. As mentioned in the previous section, failure to identify any associations between high MutPred-scoring variants (assumed to be mildly deleterious) and disease phenotypes or indicators of oxidative stress and inflammation gave rise to the fundamental biological question of whether these variants do indeed play a role in mitochondrial function. To test this, cybrid cell lines produced for a previous study were utilised, as appropriate blood samples from SABPA participants were not available. As discussed in Chapter 7, this is not problematic as the process by which cybrid cell lines are made isolates the mitochondrial component from the donor phenotype (Wilkins et al., 2014), and places it in the uniform background of the host cell line used for all cybrids (143B osteosarcoma cells in this case). Eight cell lines were identified that best met the criteria set for selection, and cultured past 35 passages in an attempt to allow RMCN to stabilise. However, even at this advanced number of passages, RMCN measurements did not stabilise and so mitochondrial respiration of the cell lines was measured using the Seahorse XF^e analyser. As a representation of mitochondrial respiration, only basal respiration rates were used in statistical analyses. The cybrid cell line SA19 showed irregular measurements of RMCN and basal respiration. These did not correspond to previous measurements of this cell line, and it was excluded from further analyses. Pearson's correlation tests between basal respiration rates of cybrids cell lines and several genetic parameters revealed that the number of variants with MutPred scores above 0.5, as well as the mutational loads calculated using these variants, could account for 17% and 16% of variation respectively. Considering that all of these variants have MutPred scores no higher than 0.72, and could be considered *mildly* deleterious, an effect size of around -.4 (Pearson's correlation coefficient) for both these parameters is considerable. Correlations of these two parameters with basal respiration rates were in a negative direction (higher MutPred loads or number of variants equals lower basal respiration rates), which makes biological sense in a straightforward simplistic manner. When cybrid cell lines were then grouped according to the number of variants with MutPred scores above 0.5 (having either zero, one, or more than two such variants), the picture became more complicated: average respiration rates for groups having zero or more than two high-scoring variants did not significantly differ from each other, and were lowest, but did significantly differ from the high average basal respiration rates of the group of cybrid cell lines having only one such variant. However, it is likely that the low basal respiration rates of SA1, having no variants with MutPred scores above 0.5, was due to lowered coupling efficiency of the cells. With the exception of this cell line, it could be that mitochondrial respiration is decreased when several high MutPred-scoring variants are present, but not when only one such variant is present.

Considering the small number of samples in this study, it is impossible to draw confident conclusions from the data before studies are repeated using a wider range of samples.

Interestingly, both RMCN and the number of variants with MutPred scores below 0.5 (which are assumed to be of little functional or structural impact) correlated significantly better with basal respiration than genetic parameters which consider only high-scoring variants. Also, these correlations were in a positive direction meaning increased RMCN or number of low MutPred-scoring variants results in increased respiration rates. For RMCN, this again can be explained by a straightforward biological model: higher RMCN suggests increased rates of translation and transcription of mitochondrial proteins, leading to increased availability of these sub-units for OXPHOS complex assembly and ultimately, higher respiration rates. Knowing whether this is indeed the case would require further elaborated studies, where the levels of intermediate compounds (such as transcripts of genes for OXPHOS subunits) in this model are also measured. The positive correlation between basal respiration rates and the number of variants with MutPred scores below 0.5 were also reflected in ANOVAS comparing groups containing cybrid cell lines which carry either four, five or six such variants. These groups all differed significantly from each other and average basal respiration rates were progressively higher in groups with five and six of these low-scoring variants. Since the number of variants called against the rCRS is affected by the position of that sequence in relation to the rCRS, the differences in the number of variants between cybrid cell lines were re-calculated using the reconstructed sapiens reference sequence (RSRS). The RSRS roots the phylogeny in such a way that the distance between all modern human sequences and the RSRS is the same (Behar et al., 2012). As discussed in Chapter 7, this step did not change the groupings or the continuous order of these cell lines, most likely because cell lines of comparable haplogroups were selected. Unlike variants with MutPred scores above 0.5, these low-scoring variants are not expected to have any direct structural or functional impact on mitochondrial enzyme proteins. As such, their influence on basal respiration rates is likely to be indirect, perhaps being involved in regulation or cross-talk with nuclear genes, the mechanisms of which are still unclear (Picard et al., 2016; Reinecke et al., 2009). One possible mechanism could be via small mitochondrial peptides, encoded by mtDNA. More than a decade ago such a peptide, called humanin, was discovered to be encoded within the mtDNA 16S rRNA coding region (Hashimoto et al., 2001). Later studies revealed humanin to be a neuroprotective factor, with circulating levels decreasing with age (Muzumdar et al., 2009). Consequently, humanin has been a target of many studies aiming to investigate the role of mitochondria in aging and age-related diseases. Recent studies have identified several other potential mitochondrial derived peptides, by searching within the 12S and 16S rRNA coding regions. Lee et al. (2016) identified a 16-amino-acid peptide named MOTS-c to be implicated in obesity and insulin resistance via regulation of the folate-methionine cycle and purine biosynthesis (and the eventual activation of AMP-activated protein kinase). Cobb et al. (2016) found six more potential small humanin like peptides (SHLPs), the codes of which lie within the mtDNA 16S rRNA

coding region. Two of these, SHLP2 and SHLP3, were shown to enhance cell viability, decrease apoptosis and ROS formation, increase respiration and ATP production *in vitro*, and to be involved in several other signalling pathways related to insulin sensitivity and inflammatory profiles. Both SHLP2 and SHLP3 also decreased with age in mice, which again might implicate these small peptides in age-related diseases. In the study by Cobb et al. (2016) only the 16S rRNA coding region was scanned for SHLPs; great potential therefore exists to discover dozens more when the rest of the mtDNA genome is investigated. How many of these mitochondrial peptides exist, how they influence mitochondrial and cellular function, and how mtDNA variants impact those functions are all questions to which the answers might also help to understand the role of mtDNA variants in disease. Some variants that currently have low MutPred pathogenicity scores and are therefore thought not to be functionally relevant, might prove to have significant consequences in the translation, transcription and function of potentially existing SHLPs. This very young branch of mitochondrial genetics certainly promises to bring some interesting perspectives and discussions to the field.

When considering drawing conclusions from this part of this study, several limitations should be noted. As discussed in Chapter 7, the Seahorse XF^{e} analyser itself imposes various technical limitations on studies, because of inter- and intra-assay variation. The limitation on the number of cell lines that can be analysed in a single run is then carried over to statistical power, rendering confident findings difficult. The availability of cybrid cell lines which do or do not meet specific selection criteria was also limiting, as not many cell lines with very rare or very high MutPred-scoring variants could be used. RMCN stability was also not achieved even after the recommended number of passages, which potentially influenced mitochondrial respiration analyses significantly. Also, other factors such as gene product transcription, ROS production, cell viability and DNA damage that could help to elucidate the underlying mechanisms involved in the impact of mtDNA variants in basal respiration were not measured in this preliminary study. Addressing these limitations is likely be the subject of future projects. As such, the objective, which was to investigate the role of mtDNA variants in mitochondrial respiration, was only partially met, given the limitations listed above.

8.4 FINAL CONCLUSIONS AND FUTURE PROSPECTS

To conclude, the aim of this study was to determine whether mtDNA variants play a significant role in the presentation or severity of cardiometabolic disease in a bi-ethnic cohort that was available for this purpose. This was done using several different approaches, that included a novel and more appropriate model: the MutPred adjusted load hypothesis.

Given the statistical power of the main study, conclusions could be drawn from the results presented in Chapters 5 and 6, where the role of non-synonymous protein coding mtDNA variants were investigated, with reasonable confidence, these were: *it is unlikely that mtDNA variants, represented by the MutPred adjusted load, play a significant role in hypertension, diabetes or markers of oxidative stress and inflammation in this cohort.*

It should be noted that the MutPred mutational load hypothesis is a flexible concept that can be modified according to the needs of a study. When SABPA participants were used, two adjustments were made to the total MutPred mutational load, the first was to exclude variants with scores below 0.5, and the second was to adjust for the phylogenetic position of a sequence in relation to that of the rCRS. When cybrid cell lines were used, this second adjustment step was not applied since population stratification was not a concern (cell lines were of comparable haplogroups). As the MutPred mutational load hypothesis is a new approach, it is likely to undergo some more modifications as its use is explored in other phenotypes and cohorts. From the limited results obtained through the cybrid cell line study, it is likely that the variants to be included in a mutational load calculation should be reconsidered. Perhaps variants with MutPred scores above 0.5 but that are known to be common haplogroup-defining variants, could be excluded, or rare variants with MutPred scores below 0.5 could be included. Evaluating a wide-ranging number of options would no doubt be helpful in the process of unravelling the effect of mtDNA in this useful, but imperfect, *in vitro* system. Another factor to reconsider is the complete exclusion of low-scoring variants from the mutational load calculations, as these variants might prove to be biologically relevant in some way. However, in this one should exercise some caution, as the relationship between low-scoring variants and basal respiration was in the opposite direction to that of high-scoring variants and basal respiration. This might indicate that these classes of variant impact on mitochondrial function in different ways. Combining all these variants into a single mutational load might mask real relationships. The more pronounced influence of population stratification on low-scoring variants can also not be ignored: if the number of variants with MutPred scores below 0.5 were investigated within the SABPA cohort, a clear relationship between this parameter and hypertension would have been found. This is simply because haplogroup L lineages have more such variants that haplogroup MN lineages, and in this cohort, haplogroup L participants had higher blood pressure than their haplogroup MN counterparts. However, as discussed in detail throughout this thesis, both these parameters are heavily influenced by non-desease factors, and such an analysis would likely result in a type I error. Thus, while the MutPred mutational load hypothesis is open to scrutiny and modification, this should be done after careful consideration. Nevertheless, it is this author's opinion that this approach delivers comparatively well-powered results, and could be a useful alternative to current approaches such as haplogroup association studies.

From this study, several potential future prospects can be identified. As discussed in Section 8.3.2, the re-evaluation of the evidence for pathogenicity of clinically proven mutations will be useful and might contribute in a positive way to diagnostic and genetic counselling fields. As mentioned throughout this thesis, the high number of rare variants found in African sequences indicates that genetic databases have not yet reported all existing mtDNA variants, and that populations from SSA are under-represented in these databases. Thus, large-scale sequencing of people from African lineages should be prioritised/supported if we are to provide population-relevant care to all nations. In relation to this, the lack of well-phenotyped studies in African populations, as well as the lack of clinical gender- and population-specific reference values complicate the design of studies in these populations, and should be addressed. It might also be useful to repeat the current study in a severely diseased cohort, or to consider the influence of mtDNA variants in disease outcomes or endpoints. A role for mtDNA copy number in diabetes complications has previously been reported (Czajka et al., 2015; Malik et al., 2015), and RMCN was significantly correlated with mitochondrial respiration in the current study (cybrids sub-study). Determining the mtDNA copy number in SABPA participants or another CVD cohort could contribute insights to this field of mtDNA genetics. Lastly, the cybrid cell lines study included in this thesis inspired many new questions and considerations. To answer these questions, this study should be repeated where the limitations listed in the previous section are addressed wherever possible. It would be advantageous to also include measurements for cell viability, ROS production and mtDNA gene expression, so that the consequences of altered mitochondrial respiration in cells can be assessed. Furthermore, repeating the cybrid evaluations using several other haplogroups would contribute greatly to the understanding of the role of both rare and common variants in mitochondrial function, and ultimately in human health and disease.

References

Achilli, A., Olivieri, A., Pala, M., Hooshiar Kashani, B. C., Perego, U. A., Gandini, F., . . . Torroni, A. (2011). Mitochondrial DNA backgrounds might modulate diabetes complications rather than T2DM as a whole. *PloSone*, 6(6): e21029.

Al-Kafaji, G., & Golbahar, J. (2013). High glucose-induced oxidative stress increases the copy number of mitochondrial DNA in human mesangial cells. *BioMed Res Int*, 2013: 754946.

Allen, J. F. (1992). Protein phosphorylation in regulation of photosynthesis. *Biochim Biophys Acta*, 1098: 275–335.

Allen, J. F. (2015). Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *PNAS*, 22: 10231–10238.

American Diabetes Association. (2010). Standards of Medical Care in Diabetes. *Diabetes Care*, 33(Suppl 1), S11–S61.

an der Walt, J., Nicodemus, K., Martin, E., Scott, W., Nance, M., Watts, R., . . . Vance, J. (2003). Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. *Am J Hum Genet*, (4): 804-811.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA]. *Nature Genetics*, 23: 147.

Ballana, E., Morales, E., Rabionet, R., Montserrat, B., Ventayol, M., Bravo, O., . . . Estivill, X. (2006). Mitochondrial 12S rRNA gene mutations affect RNA secondary structure and lead to variable penetrance in hearing impairment. *Biochem Biophys Res Commun*, 34: 950-957.

Bandelt, H.-J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. . *Mol Biol Evol*, 16:37-48.

Bandelt, H.-J., Kloss-Brandstätter, A., Richards, M. Y.-G., & Logan, I. (2013). The case for the continuing use of the revised Cambridge reference sequence (rCRS) and the stadardization of notation in human mitochondrial DNA studies. *J Hum Genet*, 59: 66-77.

Baudouin, S., Saunders, D., Tiangyou, W., Elson, J., Poynter, J., Pyle, A., . . . Chinnery, P. (2005). Mitochondrial DNA and survival after sepsis: a prospective study. *Lancet*, 366: 2118-2121.

Behar, D., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E., Silva, N., . . . Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet.*, 90: 675-84.

Bruno, C., Sacco, O., Santorelli, F., Assereto, S., Tonoli, E., Bado, M., . . . Minetti, C. (2003). Mitochondrial myopathy and respiratory failure associated with a new mutation in the mitochondrial transfer ribonucleic acid glutamic acid gene. *J Child Neurol.*, 18:300-3.

Canter, J., Kallianpur, A., Parl, F., & Millikan, R. (2005). Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Research*, 65: 8028-8033.

Cardena, M., Ribeiro-Dos-Santos, A., Santos, S., Mansur, A., Bernardez-Pereira, S., Santos, P., . . . Fridman, C. (2016). Mitochondrial and genomic ancestry are associated with etiology of heart failure in Brazilian patients. *J. Hum. Hypertens.*, 30(2):120-3.

Cardena, M., Ribeiro-dos-Santos, A., Santos, S., Mansur, A., Pereira, A., & Fridman, C. (2014). Amerindian genetic ancestry is associated with higher survival rates compared to African and European ancestry in Brazilian patients with heart failure. *Int. J. Cardiol.*, 176(2):527-8.

Castro, J., El-Atat, F., McFarlane, S., Aneja, A., & Sowers, J. (2003). Cardiometabolic syndrome: pathophysiology and treatment. *Curr Hypertens Rep*, 5:393-401.

Cavadas, B., Soares, P., Camacho, R., Brandao, A., Costa, M. D., Fernandes, V., . . . Periera, L. (2015). Fine time scaling of purifying selection on human nonsynonymous mtDNA mutations based on the worldwide population tree and mother-child pairs. *Human mutation*, 36(11): 1100-1111.

Chance, B., Sies, H., & Boveris, A. (1979). Hydroperoxide metabolism in mammalian organs. *Physiological Reviews*, 59: 527–605.

Chinnery, P. F., Elliott, H. R., Syed, A., Rothwell, P. M., & Study., O. V. (2010). Mitochondrial DNA haplogroups and risk of transient ischaemic attack and ischaemic stroke: a genetic association study. *Lancet Neurology*, 9(5):498-503.

Chomyn, A. (1996). Platelet-mediated transformation of human mitochondrial DNA-less cells. *Methods Enzymol*, 264: 334-339.

Cobb, L. J., Lee, C., Xiao, J., Yen, K., Wong, R. G., Nakamura, H. K., . . . Cohen, P. (2016). Naturally occurring mitochondrial-derived peptides are agedependent regulators of apoptosis, insulin sensitivity, and inflammatory markers. *Aging*, 8: 796-809.

Czajka, A., Ajaz, S., Gnudi, L., Parsade, C., Jones, P., Reid, F., & Malik, A. (2015). Altered mitochondrial function, mitochondrial DNA and reduced metabolic flexibility in patients with diabetic nephropathy. *EBioMedicine*, 2: 499-512.

de Champlain, J., Wu, R., Girouard, H., Karas, M., EL Midaoui, A., Laplante, M., & Wu, L. (2004). Oxidative stress in hypertension. *Clin Exp Hypertens*, 26(7-8):593-601.

DeBalsi, K. L., Hoff, K. E., & Copeland, W. C. (2016). Role of the mitochondrial DNA replication machinery in mitochondrial DNA mutagenesis, aging and age-related diseases. *Ageing Research Reviews*.

DiMauro, S., & Schon, E. (2001). Mitochondrial DNA mutations in human disease. *Am J Med Genet*, 106:18–26.

Dranka, B. P., Benavides, G. A., Diers, A. R., Giordano, S., Zelickson, B. R., Reily, C., . . . Darley-Usmar, V. M. (2011). Assessing bioenergetic function in response to oxidative stress by metabolic profiling. *Free Radic Biol Med*, 51: 1621–1635. .

Elson, J. L., Herrnstadt, C., Preston, G., Thal, L., Morris, C. M., Edwardson, J. A., . . . Howell, N. (2006). Does the mitochondrial genome play a role in the etiology of Alzheimer's disease? *Human Genetics*, 119: 241-254.

Elson, J., & Lightowlers, R. (2006). Mitochondrial DNA clonality in the dock: can surveillance swing the case? *Trends Genet.*, 22:603-607.

Elson, J., Smith, P., Greaves, L., Lightowlers, R., Chrzanowska-Lightowlers, Z., Taylor, R., & Vila-Sanjurjo, A. (2015). The presence of highly disruptive 16S rRNA mutations in clinical

samples indicates a wider role for mutations of the mitochondrial ribosome in human disease. *Mitochondrion*, 25: 17-27.

Elson, J., Turnbull, D., & Howell, N. (2004). Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *American Journal of Human Genetics*, 74: 229-238.

Fisher-Wellman, K. H., & Neufer, P. D. (2012). Linking mitochondrial bioenergetics to insulin resistance via redox biology. *Trends in Endocrinol and Metab*, 23: 142-153.

Florentz, C., Sohm, B., Tryoen-Tóth, P., Pütz, J., & Sissler, M. (2003). Human mitochondrial tRNAs in health and disease. *Cell Mol Life Sci*, 60:1356-1375.

Gómez-Durán, A., Pacheu-Grau, D., López-Gallardo, E., . . . E. (2010). Unmasking the causes of multifactorial disorders: OXPHOS differences between mitochondrial haplogroups. *Hum Mol Genet*, 19: 3343-3353.

Guan, M., Fischel-Ghodsian, N., & Attardi, G. (1996). Biochemical evidence for nuclear gene involvement in phenotype of non-syndromic deafness associated with mitochondrial 12S rRNA mutation. *Hum Mol Genet*, 5:963–971.

Guan, M., Fischel-Ghodsian, N., & Attardi, G. (2000). A biochemical basis for the inherited susceptibility to aminoglycoside ototoxicity. *Hum Mol Genet*, 9: 1787-1793.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., . . . Kamali, A. (2015). The African genome variation project shapes medical genetics in Africa. *Nature*, 517: 327-332.

Hamer, M., Malan, L., Malan, N., Schutte, A., Huisman, H., van Rooyen, J., . . . Seedat, Y. (2011). Objectively assessed health behaviors and sub-clinical atherosclerosis in black and white Africans: The SABPA study. *Atherosclerosis*, 215: 237-242.

Hamer, M., von Känel, R., Reimann, M., Malan, N. T., Schutte, A. E., Huisman, H. W., & Malan, L. (2015). Progression of cardiovascular risk factors in black Africans: 3 year follow up of the SABPA cohort study. *Atherosclerosis*, 238(1), 52–54.

Harrison, D. G., Gongora, M. C., Guzik, T. J., & Widder, J. (2007). Oxidative stress and hypertension. *Journal of the American Society of Hypertension*, 1(1): 30-44.

Hashimoto, Y., Niikura, T., Tajima, H., Yasukawa, T., Sudo, H., Ito, Y., . . . Nishimoto, I. (2001). A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and Abeta. *Proc Natl Acad Sci U S A*, 98:6336-6341.

Hastedt, M., Büchner, M., Rothe, M., Gapert, R., Herre, S., Krumbiegel, F., . . . Hartwig, S. (2013). Detecting alcohol abuse: traditional blood alcohol markers compared to ethyl glucuronide (EtG) and fatty acid ethyl esters (FAEEs) measurement in hair. *Forensic Sci Med Pathol*, 9: 471-477.

Hayashi, I., Morishita, Y., Imai, K., Nakamura, M., Nakachi, K., & Hayashi, T. (2007). High-throughput spectrophotometric assay of reactive oxygen species in serum. *Mut Res*, 631: 55–61.

Hermida, R. C., Ayala, D. E., Mojon, A., Fontao, M. J., & Chayan, L. F. (2013). Differences between men and woman in Ambulatory blood pressure thresholds for diagnosis of hypertension based on cardiovascular outcomes. *Chronobiology international*, 30(1-2): 221-232.

Hoebel, S., Malan, L., & De Ridder, J. (2013). Determining ethnic-, gender-, and age-specific waist circumference cut-off points to predict metabolic syndrome: the Sympathetic Activity and Ambulatory Blood Pressure in Africans (SABPA) study. *JEMDSA*, 18: 88-96.

Howell, N., Elson, J., PF, C., & Turnbull, D. (2005). MtDNA mutations and common neurodegenerative disorders. *Trends Genet*, 21: 583-586.

Hu, C., Wu, M., Chao, M., Pan, C., Wang, C., Swenberg, J., & Wu, K. (2004). Comparison of analyses of urinary 8-hydroxy-2'- deoxyguanosine by isotope-dilution liquid chromatography with electrospray tandem mass spectrometry and by enzymelinked immunosorbent assay. *Rapid Commun Mass Spectrom*, 18:505–510.

Hudson, G., Carelli, V., Spruijt, L., Gerards, M., Mowbray, C., Achilli, A., . . . Chinnery, P. F. (2007). Clinical expression of Leber hereditary optic neuropathy is affected by the mitochondrial DNA-haplogroup background. *American journal of human genetics*, 81(2): 228-233.

Hudson, G., Gomez-Duran, A., Wilson, I. J., & Chinnery, P. F. (2014). Recent Mitochondrial DNA Mutations Increase the Risk of Developing Common Late-Onset Human Diseases. *PLoS Genet* , 10(5): e1004369.

Huisman, H. W., Schutte, R., Schutte, A. E., Rooyen, J. M., Malan, N. T., Fourie, C. M., . . . Malan, L. (2012). The usefulness of γ-glutamyltransferase as a marker of cardiovascular function in Africans and Caucasians: The SABPA study. *Clinical and Experimental Hypertension*, 34: 8–16.

Janero, D. (1990). Malondialdehyde and thiobarbituric acid-reactivity as diagnostic indices of lipid peroxidation and peroxidative tissue injury. *Free Radic Biol Med*, 9: 515–540.

Jansen van Vuren, E., Malan, L., Cockeran, M., Scheepers, J., Oosthuizen, W., & Malan, N. (2016a). Fibrosis and coronary perfusion - a cardiovascular disease risk in an African male cohort: The SABPA study. *Clin Exp Hypertens*, 38: 482-488.

Jansen van Vuren, E., Malan, L., von Känel, R., Cockeran, M., & Malan, N. (2016b). Hyperpulsatile pressure, systemic inflammation and cardiac stress are associated with cardiac wall remodeling in an African male cohort: the SABPA study. *Hypertension Research*, 39, 648-653.

Janssen, G. M., Hensbergen, P. J., Bussel, F. J., Balog, C. I., Maassen, J. A., Deelder, A. M., & Raap, A. K. (2007). The A3243G tRNALeu(UUR) mutation induces mitochondrial dysfunction and variable disease expression without dominant negative acting translational defects in complex IV subunits at UUR codons. *Hum Mol Genet*, 16: 2472-2481.

Ji, Y., Zhang, A., Jia, X., Zhang, Y., Xiao, X., Li, S., . . . Yao, Y. (2008). Mitochondrial DNA haplogroups M7b1'2 and M8a affect clinical expression of Leber hereditary optic neuropathy in Chinese families with m.11778G>A mutation. *Am. J. Hum. Gen.*, 83(6): 760-768.

Johnston, I. G., & Williams, B. P. (2016). Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retension. *Cell Systems*, 2: 101–111.

Kenney, M. C., Chwa, M., Atilano, S. R., Falatoonzadeh, P., Ramirez, C., Malik, D., . . . Udar, N. (2014a). Molecular and Bioenergetic Differences between Cells with African versus European Inherited mitochondrial DNA haplogroups: implications for population susceptibility to diseases. *Biochim Biophys Acta*, 1842: 208–219.

Kenney, M. C., Chwa, M., Atilano, S. R., Falatoonzadeh, P., Ramirez, C., Malik, D., . . . Udar, N. (2014b). Inherited mitochondrial DNA variants can affect complement, inflammation and

apoptosis pathways: insights into mitochondrial–nuclear interactions. *Human Molecular Genetics*, 23: 3537–3551.

Kenney, M., Falatoonzadeh, P., Atilano, S., Chwa, M., Caceres-del-Carpio, J., Malik, D., . . . Kuppermann, B. (2016). African-origin mitochondrial DNA variants as a contributing factor to susceptibilities for diabetes and age-related diseases. *Int J Diabetes Clin Res*, 3:053.

Kern, A., & Kondrashov, F. (2004). Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. *Nat Genet*, 36:1207-1212.

Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K., . . . Oefner, P. J. (2006). The Role of Selection in the Evolution of Human Mitochondrial Genomes. *Genetics*, 172: 373–387.

Kruger, R., Schutte, R., Huisman, H., Rooyen, J. V., Malan, N., Fourie, C., . . . Schutte, A. (2012). Associations between reactive oxygen species, blood pressure and arterial stiffness in black South Africans: the SABPA study. *Journal of Human Hypertension*, 26: 91–97.

Lee, C., Zeng, J., Drew, B. G., Sallam, T., Martin-Montalvo, A., Wan, J., . . . Cohen, P. (2015). The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metabolism* , 21: 443–454.

Lee, H., Song, J., Shin, C., Park, D., Park, K., Lee, K., & Koh, C. (1998). Decreased mitochondrial DNA content in peripheral blood precedes the development of non-insulin-dependent diabetes mellitus. *Diabetes Res Clin Pract*, 42: 161–167.

Lee, J. E., Park, H., Ju, Y. S., Kwak, M., Kim, J.-I., Oh, H. Y., & Seo, J.-S. (2009). Higher mitochondrial DNA copy number is associated with lower prevalence of microalbuminuria. *Experimental & Molecular Medicine*, 41: 253-258.

Li, B., Krishnan, V. G., Mort, M. E., Xin, F., Kamati, K. K., Cooper, D. N., . . . Radivojac, P. (2009). Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*, 25(21): 2744-2750.

Liu, X., Kim, C., Yang, J., Jemmerson, R., & Wang, X. (1996). Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome c. *Cell*, 86:147-57.

Lui, C., Yang, Q., Hwang, S., Sun, F., Johnson, A. D., Shirihai, O. S., . . . Schwartz, F. (2012). Association of geneic variation in the mitochondrial genome with blood pressure and metabolic traits. *Hypertension*, 60: 949-956.

Maeda, K., Kawai, H., Sanada, M., . . . Majima, H. (2016). Clinical Phenotype and Segregation of Mitochondrial 3243A>G Mutation in 2 Pairs of Monozygotic Twins. *JAMA*.

Maier, U., Zauner, S., Woehle, C., Bolte, K., Hempel, F., Allen, J., & and Martin, W. (2013). Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes. *Genome Biol Evol*, 5: 2318–2329.

Malan, L., & Malan, N. (2016). Emotional Stress as a Risk for Hypertension in Sub-Saharan Africans: Are We Ignoring the Odds? *Adv Exp Med Biol*, doi: 10.1007/5584-2016-37.

Malan, L., Hamer, M., Frasure-Smith, N., Steyn, H. S., & Malan, N. T. (2015). Cohort Profile: Sympathetic activity and Ambulatory Blood Pressure in Africans (SABPA) prospective cohort study. *International Journal of Epidemiology*, 44(6):1814-1822.

Malan, L., Hamer, M., Schlaich, M., Lambert, G., Ziemssen, T., Reimann, M., . . . Malan, N. (2013a). Defensive active coping facilitates chronic hyperglycemia and endothelial dysfunction in African men: the SABPA study. *Int J Cardiol*, 168: 999-1005.

Malan, L., Hamer, M., Schlaich, M., Lambert, G., Ziemssen, T., Reimann, M., . . . Malan, N. (2013b). Defensive coping facilitates higher blood pressure and early sub-clinical structural vascular disease via alterations in heart rate variability: the SABPA study. *Atherosclerosis*, 227: 391-397.

Malan, L., Malan, N., Wissing, M., & Seedat, Y. (2008). Coping with urbanization: a cardiometabolic risk? The THUSA study. *Biol Psychol*, 79: 323-328.

Malan, L., Schutte, C., Alkerwi, A., Stranges, S., & Malan, N. (2017). Hypothalamic-pituitary-adrenal-axis dysregulation and double product increases potentiate ischemic heart disease risk in a Black male cohort: The SABPA study. *Hypertens Research*, DOI:10.1038/hr.2017.5 .

Malik, A., Parsade, C., Ajaz, S., Crosby-Nwaobi, R., Gnudi, L., Czajka, A., & Sivaprasad, S. (2015). Altered circulating mitochondrial DNA in and increased inflammation in patients with diabetic retinopathy. *Diabetic research and clinical practice*, 110: 257-265.

Mancia, G., Fagard, R., Narkiewicz, K., Redón, J., Zanchetti, A., Böhm, M., . . . Ruilope, L. M. (2013). 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Journal of hypertension*, 31(7):1281-357.

McFarland, R., Elson, J., Taylor, R.W., Howell, N., & Turnbull, D. (2004). Assigning pathogenicity to mitochondrial tRNA mutations: when ‘‘definitely maybe’’ is not good enough. *Trends Genet*, 20:591–596.

McKenzie, M., & Duchen, M. (2016). Impaired Cellular Bioenergetics Causes Mitochondrial Calcium Handling Defects in MT-ND5 Mutant Cybrids. *PLoS ONE* , 11: e0154371.

McKenzie, M., Liolitsa, D., Akinshina, N., Campanella, M., Sisodiya, S., Hargreaves, I., . . . Duchen, M. (2007). Mitochondrial ND5 gene variation associated with encephalomyopathy and mitochondrial ATP consumption. *J Biol Chem*, 82: 36845–36852.

Meissner-Roloff, M. (2009). *The occurrence of mitochondrial DNA polymerase gamma gene mutations in mitochondrial deficiencies, in a selection of South African paediatric patients.* Potchefstroom: North-West University (Dissertation - Masters). http://hdl.handle.net/10394/5095 .

Meldau, S., Riordan, G., Van der Westhuizen, F., Elson, J., Smuts, I., Pepper, M., & Soodyal, l. H. (2016). Could we offer mitochondrial donation or similar assisted reproductive technology to South African patients with mitochondrial DNA disease? *S Afr Med J*, 106:234-236.

Mels, C., Schutte, A., Schutte, R., Pretorius, P., Smith, W., Huisman, H., . . . L, M. (2014). 8-Oxo-7,8-dihydro-2’-deoxyguanosine, reactive oxygen species and ambulatory blood pressure in African and Caucasian men: The SABPA study. *Free Radic Res*, 48:1291-9.

Mensah, G. A. (2013). Descriptive Epidemiology of Cardiovascular Risk Factors and diabetes in Africa. *Progress in Cardiovascular diseases*, 56: 240-250.

Mensah, G. A., Roth, G. A., Sampson, U. K., Moran, A. E., Feigin, V. L., Forouzanfar, M. H., . . . Murray, C. J. (2015). Mortality from cardiovascular diseases in sub-Saharan Africa, 1990–2013: a systematic analysis of data from the Global Burden of Disease Study 2013. *Cardiovasc J Afr*, 26(2 H3Africa Suppl): s6–s10. .

Mensah, G., Peprah, E., Sampson, U., & Cooper, R. (2015). H3Africa comes of age. *Cardiovasc J Afr*, 26(2 H3Africa Suppl): s3–s5.

Mereschkowsky, K. (1910). Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Ent-stehung der Organismen. *Biol Centralbl*, 30: 353-367.

Mitchell, A. L., Elson, J. L., Howell, N., Taylor, R. W., & Turnbull, D. M. (2006). Sequence variation in mitochondrial complex I genes: mutation or polymorphism? *Journal of Medical Genetics*, 43(2):175-9.

Mitra, K., Wunder, C., Roysam, B., Lin, G., & Lippincott-Schwartz. (2009). A hyperfused mitochondrial state achieved at G1-S regulates cyclin E buildup and entry into S phase. *PNAS*, 106: 11960-11965.

Mokhaneli, M., F. C., Botha, S., & Mels, C. (2016). The association of oxidative stress with arterial compliance and vascular resistance in a bi-ethnic population: the SABPA study. *Free Radic Res*, 50: 920-928.

Moran, A., Forouzanfar, M., Sampsonc, U., Chugh, S., Feigine, V., & Mensah, G. A. (2013). The epidemiology of cardiovascular diseases in Sub-Saharan Africa: The global burden of diseases, Injuries and risk factors 2010 study. *Progress in cardiovascular disease*, 56: 234-239.

Mosteller, R. (1987). Simplified calculation of body surface area. *New Eng J Med*, 317:1098.

Muzumdar, R., Huffman, D., Atzmon, G., Buettner, C., Cobb, L., Fishman, S., . . . Cohen, P. (2009). Humanin: a novel central regulator of peripheral insulin action. *PLoS One*, 4:e6334.

Nakahira, K., Haspel, J., Rathinam, V., Lee, S., Dolinay, T., Lam, H., . . . Choi, A. (2011). Autophagy proteins regulate innate immune responses by inhibiting the release of mitochondrial DNA mediated by the NALP3 inflammasome. *Nat Immunol.*, 12(3):222-30.

Nunnari, J., & Suomalainen, A. (2012). Mitochondria: in sickness and in health. *Cell*, 148:1145-159.

Oka, T., Hikoso, S., Yamaguchi, O., Taneike, M., Takeda, T., Tamai, T., . . . Otsu, K. (2012). Mitochondrial DNA that escapes from autophagy causes inflammation and heart failure. *Nature*, 485(7397): 251-5.

Okin, P. M., Kjeldsen, S. E., Dahlo¨f, B., & Devereux, R. B. (2011). Racial Differences in Incident Heart Failure During antihypertensive therapy. *Circulation and Cardiovascular Qualitative Outcomes*, 4:157-164.

Omboni, S., Aristizabal, D., De La Sierra, A., Dolan, E., G, H., Kahan, T., . . . Parati. (2016). Hypertension types defined by clinic and ambulatory blood pressure in 14 143 patients referred to hypertension clinics worldwide. Data from the ARTEMIS study. *J Hypertens*, 34:2187-2198.

Otaegui, D., Paisán, C., Sáenz, A., Martí, I., Ribate, M., Martí-Massó, J., . . . López de Munain, A. (2004). Mitochondrial polymporphisms in Parkinson's Disease. *Neurosci Lett*, 370:171-4.

Owolabi, M., Arulogun, O., Melikam, S., Adeoye, A., Akarolo-Anthony, S., Akinyemi, R., . . . Owolabi, L. (2015). The burden of stroke in Africa: a glance at the present and a glimpse into the future. *Cardiovasc J Afr*, 26(2 H3Africa Suppl): S27–S38.

Pepys, M., & Hirschfield, G. (2003). C-reactive protein: a critical update. *The Journal of Clinical Investigation*, 111: 1805–1812.

Pereira, L., Soares, P., Radivoiac, P., Li, B., & Samuels, D. C. (2011). Comparing phylogeny and the predicted pathogenicity of protein variations reveals equal purifying selection across the global human mtDNA diversity. *American Journal of Human Genetics*, 8;88(4):433-9.

Picard, M., Wallace, D. C., & Burelle, Y. (2016). The rise of mitochondria in medicine. *Mitochondrion*, 30: 105-116.

Piepoli, M., Hoes, A., Agewall, S., Albus, C., Brotons, C., Catapano, A., . . . Smulders. (2016). European Guidelines on cardiovascular disease prevention in clinical practice. The Sixth Joint Force of the European Society of Cardiology and other Societies on cardiovascular diasease prevention in clinical practice. *Eur Heart J*, 5:1-78.

Qin, Y., Xue, L., Jiang, P., Xu, M., He, Y., Shi, S., . . . Guan, M.-X. (2014). Mitochondrial tRNA variants in Chinese subjects with coronary heart disease. *Journal of the American Heart Association*, 3: e000437.

Raimundo, N., Song, L., Shutt, T., McKay, S., Cotney, J., Guan, M., . . . Shadel, G. (2012). Mitochondrial stress engages E2F1 apoptotic signaling to cause deafness. *Cell*, 148: 716-26.

Reimann, M., Hamer, M., Malan, N., Schlaich, M., Lambert, G., Ziemssen, T., . . . Malan, L. (2013). Effects of acute and chronic stress on the L-arginine nitric oxide pathway in black and white South Africans: the sympathetic activity and ambulatory blood pressure in Africans study. *Psychosom Med.*, 75: 751-758.

Reinecke, F., Smeitink, J., & van der Westhuizen, F. (2009). OHPHOS gene expression and control in mitochondrial disorders. *Biochem Biophys Acta*, 1792: 1113-1121.

Rossignol, R., Gilkerson, R., Aggeler, R., Yamagata, K., Remington, S. J., & Capaldi, R. A. (2004). Energy substrate modulates mitochondrial structure and oxidative capacity in cancer cells. *Cancer res*, 64: 985–993.

Salas, A., & Elson, J. L. (2015). Mitochondrial DNA as a risk factor for false positives in case-control association studies. *Journal of Genetics and Genomics*, Apr 20;42(4):169-72.

Samuels, D. (2005). Life span is related to the free energy of mitochondrial DNA. *Mech. Ageing Dev*, 126: 1123–1129.

Samuels, D. C., Carothers, A. D., Horton, R., & Chinnery, P. F. (2006). The power to detect disease associated with mitochondrial DNA haplogroups. *American Journal of Human Genetics*, 78(4):713-20.

Sanders, C. L., Lohr, K. J., Gambill, H. L., Curran, R. B., & Curran, J. F. (2008). Anticodon loop mutations perturb reading frame maintenance by the E site tRNA. *RNA*, 14: 1874–1881.

Sandhir, R., Halder, A., & Sunkaria, A. (2016). Mitochondria as a entrally positioned hub in the innate immune response. *Biochimica et Biophysica Acta*, http://dx.doi.org/10/1016/j.bbadis.2018.10.020.

Saude, E., Adamko, D., Rowe, B., Marrie, T., & Sykes, B. (2007). Variation of metabolites in normal human urine. *Metabolomics*, 3:439–451.

Schwedhelm, E., Tan-Andresen, J., Maas, R., Riederer, U., Schulze, F., & Böger, R. (2005). Liquid chromatography-tandem mass spectrometry method for the analysis of asymmetric dimethylarginine in human plasma. *Clin Chem 2005*, 1:1268Y71.

Setiawan, V., Chu, L., John, E., Ding, Y., Ingles, S., Bernstein, L., . . . Neuhausen, S. (2008). Mitochondrial DNA G10398A variant is not associated with breast cancer in African-American women . *Cancer Genetics and Cytogenetics*, 181: 16-19 .

Shimada, K., Crother, T., Karlin, J., Dagvadorj, J., Chiba, N., Chen, S., . . . Arditi, M. (2012). Oxidized mitochondrial DNA activates the NLRP3 inflammasome during apoptosis. *Immunity*, 23;36(3): 401-14.

Silva, D. F., Selfridge, J. E., Lu, J., E, L., Roy, N., Hutfles, L., . . . Swerdlow, R. H. (2013). Bioenergetic flux, mitochondrial mass and mitochondrial morphology dynamics in AD and MCI cybrid cell lines. *Human Molecular Genetics*, 22: 3931–3946.

Singh, K., Kulawiec, M., Still, I., Desouki, M., Geradts, J., & Matsui, S. (2005). Inter-genomic cross talk between mitochondria and the nucleus plays an important role in tumorigenesis. *Gene*, 354:140-146.

Singh, R., Hattersley, A., & Harries, L. (2007). Reduced peripheral blood mitochondrial DNA content is not a risk factor for Type 2 diabetes. *Diabet Med*, 24: 784–787.

Sternberg, D., Chatzoglou, E., Laforet, P., Fayet, G., Jardel, C., Blondy, P., . . . Lombes, A. (2001). Mitochondrial DNA transfer RNA gene sequence variations in patients with mitochondrial disorders. *Brain*, 124: 984-994.

Sylvester, J. E., Fischel-Ghodsian, N., Mougey, E. B., & O'Brien, T. W. (2004). Mitochondrial ribosomal proteins: Candidate genes for mitochondrial disease. *Genetics in Medicine*, 6: 73–80.

Taylor, S., Ericson, N., Burton, J., Prolla, T., Silber, J., Shendure, J., & Bielas, J. (2014). Targeted enrichment and high-resolution digital profiling of mitochondrial DNA deletions in human brain. *Aging cell*, 13: 29-38.

Trinei, M., Berniakovich, I., Pelicci, P., & Giogio, M. (2006). Mitochondrial DNA copy number is regulated by cellular proliferation: A role for Ras and p66Shc. *Biochimic Biophys Acta*, 1757: 624-630.

Tuppen, H. A., Blakely, E. L., Turnbull, D. M., & Taylor, R. W. (2010). Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta*, 1797(2):113-28.

van der Walt, E. M., Smuts, I., Taylor, R. W., Elson, J. L., Turnbull, D. M., Louw, R., & van der Westhuizen, F. H. (2012). Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease. *Journal of Human Genetics*, 20(6):650-656.

van der Westhuizen, F. H., Sinxadi, P. Z., Dandara, C., Smuts, I., Riordan, G., Meldau, S., . . . Elson, J. L. (2015). Understanding the Implications of Mitochondrial DNA Variation in the Health of Black Southern African Populations: The 2014 Workshop. *Human Mutation*, 36(5):569-71.

van Deventer, C. (2015). *Metabolomics of hypertension in South Africans: The SABPA Study*. Potchefstroom: North-West Unversity (Thesis - PhD).

van Deventer, C., Lindeque, J., van Rensburg, P., Malan, L., van der Westhuizen, F., & Louw, R. (2015). Use of metabolomics to elucidate the metabolic perturbation associated with hypertension in a black South African male cohort: the SABPA study. *J Am Soc Hypertens*, 9:104-114.

van Dyk, H. (2013). *Development of cytoplasmic hybrid cells to evaluate mitochondrial DNA mutation pathogenicity.* Potchefstroom: North-West University (Mini-dissertation - Honours).

van Dyk, H. (2016). *Evaluating the involvement of mtDNA variants in patients diagnosed with myalgic encephalomyelitis.* Potchefstroom, North-West Province, South Africa: North-West University (Dissertation - Masters). http://hdl.handle.net/10394/17668.

Venter, M., Malan, L., van Dyk, E., Elson, J. L., & van der Westhuizen, F. H. (2017). Using MutPred derived mtDNA load scores to evaluate mtDNA variation in hypertension and diabetes in a two-population cohort: The SABPA study. *Journal of genetics and genomics*, http://dx.doi.org/10.1016/j.jgg.2016.12.003.

Wajant, H., Pfizenmaier, K., & Scheurich, P. (2003). Tumor necrosis factor signaling. *Cell Death Differ*, 10: 45–65.

Wallace, D. C. (2013). A mitochondrial bioenergetic etiology of disease. *J Clin Invest*, 123: 1405–1412.

Weissensteiner, H. P., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., . . . Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucl. Acids. Res*, 44: W58-W63.

West, A., Khoury-Hanold, W., Staron, M., Tal, M., Pineda, C., Lang, S., . . . Shadel, G. (2015). Mitochondrial DNA stress primes the antiviral innate immune response. *Nature*, 520(7548): 553-7.

Wilkins, H. M., Carl, S. M., & Swerdlow, R. H. (2014). Cytoplasmic hybrid (cybrid) cell lines as a practical model for mitochondriopathies. *Redox Biology*, 2: 619–631.

World medical association. (2004). World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *J. Int. Bioethique*, 124.

Wu, L., Chiou, C., Chang, P., & Wu, J. (2004). Urinary 8-OHdG: a marker of oxidative stress to DNA and a risk factor for cancer, atherosclerosis and diabetics. *Clin Chim Acta.*, 339: 1-9.

Yarham, J. W., Al-Dosary, M., Blakely, E. L., Alston, C. L., Taylor, R. W., Elson, J. L., & McFarland, R. (2011). A comparative analysis approach to determining the pathogenicity of mitochondrial tRNA mutations. *Human Mutation*, 32(11):1319-25.

Yarham, J. W., Elson, J. L., Blakely, E. L., McFarland, R., & Taylor, R. W. (2010). Mitochondrial tRNA mutations and disease. *Wiley Interdiscip Rev RNA*, 1(2):304-24.

Yu, E., & Bennett, M. (2016). The role of mitochondrial DNA damage in the development of atherosclerosis. *Free Radic Biol Med,* 100:223-230.

Yu, E., Calvert, P., Mercer, J., Harrison, J., Baker, L., Figg, N., . . . Bennett, M. (2013). Mitochondrial DNA damage can promote atherosclerosis Independently of reactive oxygen species through effects on smooth muscle cells and monocytes and correlates with higher-risk plaques in humans. *Circulation*, 128: 702-712.

Yusuf, S., Reddy, S., Ounpuu, S., & Anand, S. (2001). Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization. *Circulation*, 104: 2746–2753.

Yu-Wai-Man, P., & Chinnery, P. (2000 [Updated 2016]). Leber hereditary optic neuropathy. In R. Pagon, M. Adam, H. Ardinger, & e. al, *GeneReviews® [Internet]*. Seattle (WA): University of Washington. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK1174/

Zhao, H., Young, W., Yan, Q., Li, R., Cao, J., Wang, Q., . . . Guan, M. (2005). Functional characterization of the mitochondrial 12S rRNA C1494T mutation associated with aminoglycoside-induced and non-syndromic hearing loss. *Nucleic Acids Res*, 33: 1132-1139.

Zhou, M.-c., Zhu, L., Cui, X., Feng, L., Zhao, X., He, S., . . . Li, Y. (2016). Reduced peripheral blood mtDNA content is associated with impaired glucose-stimulated islet β cell function in a Chinese population with different degrees of glucose tolerance. *Diabetes Metab Res Rev*, 32: 768–774.

Appendices

APPENDIX A: SABPA STUDY ETHICS APPROVAL

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

Private Bag X6001, Potchefstroom
South Africa 2520

Tel (018) 299-4900
Faks (018) 299-4910
Web http://www.nwu.ac.za

Dr L Malan

Ethics Committee
Tel +27 18 299 2542
Fax +27 18 297 5308
Email Ethics@nwu.ac.za

Dear Dr Malan

6 February 2008

ETHICS APPROVAL OF PROJECT

The North-West University Ethics Committee (NWU-EC) hereby approves your project as indicated below. This implies that the NWU-EC grants its permission that, provided the special conditions specified below are met and pending any other authorisation that may be necessary, the project may be initiated, using the ethics number below.

Project title: SABPA (Sympathetic activity and Ambulatory Blood Pressure in Africans)

Ethics number: NWU-00037-07-S6

Approval date: 12 November 2007 **Expiry date**: 11 November 2012

Special conditions of the approval (if any): None

General conditions:

While this ethics approval is subject to all declarations, undertakings and agreements incorporated and signed in the application form, please note the following:

- *The project leader (principle investigator) must report in the prescribed format to the NWU-EC:*
  - *annually (or as otherwise requested) on the progress of the project;*
  - *without any delay in case of any adverse event (or any matter that interrupts sound ethical principles) during the course of the project.*
- *The approval applies strictly to the protocol as stipulated in the application form. Would any changes to the protocol be deemed necessary during the course of the project, the project leader must apply for approval of these changes at the NWU-EC. Would there be deviation from the project protocol without the necessary approval of such changes, the ethics approval is immediately and automatically forfeited.*
- *The date of approval indicates the first date that the project may be started. Would the project have to continue after the expiry date, a new application must be made to the NWU-EC and new approval received before or on the expiry date.*
- *In the interest of ethical responsibility the NWU-EC retains the right to:*
  - *request access to any information or data at any time during the course or after completion of the project;*
  - *withdraw or postpone approval if:*
    - *any unethical principles or practices of the project are revealed or suspected;*
    - *it becomes apparent that any relevant information was withheld from the NWU-EC or that information has been false or misrepresented;*
    - *the required annual report and reporting of adverse events was not done timely and accurately;*
    - *new institutional rules, national legislation or international conventions deem it necessary.*

The Ethics Committee would like to remain at your service as scientist and researcher, and wishes you well with your project. Please do not hesitate to contact the Ethics Committee for any further enquiries or requests for assistance.

Yours sincerely

Prof M M J Lowes
(chair NWU Ethics Committee)

APPENDIX B: SABPA STUDY PARTICIPATION INFORMATION AND CONSENT FORM

NORTH-WEST UNIVERSITY

POTCHEFSTROOM CAMPUS

SCHOOL FOR PHYSIOLOGY, NUTRITION AND CONSUMER SCIENCES

PARTICIPANT INFORMATION AND CONSENT FORM

PART 1

PRINCIPAL RESEARCHER: Prof Leoné Malan, Subject Group Physiology
PROJECT LEADERS: Prof Leoné Malan

Associate Researcher(s): The postdoctoral fellow involved in this trial is Dr. P Szabolcs. Other persons assisting in the study are Proff Alta E Schutte, Hugo W. Huisman, Johannes M. van Rooyen, Nico T. Malan, Dr Rudolph Schutte, Mrs. Carla M.T. Fourie, Mrs. Tina Scholtz (Cardiovascular research group, Physiology), Proff. Hans de Ridder (Anthropometry, Physical activity), Marié Wissing, (Psychology), Linda Brand & Brian Harvey (Pharmacology), Kobus Mentz (Education), Francois van der Westhuizen (Biochemistry), Hester Klopper (Nursing), Yackoob Seedat (ECG, Kwazulu Natal), Paul Rheeder (Sonar, Pretoria University), Drs. Johan Potgieter & Mr Thumi Khumalo (Psychology), Proff Nancy Frasure-Smith & Francois Lespérance (Psychology, Canada), Drs Alaa Alkerwi (Metabolic syndrome, Luxembourg), T Ziemssen (Autonomic function, Germany), M Hamer (Inflammation, CVD, England), Prof E de Geus (Depression genetics, CVD, Netherlands).

This Participant Information and Consent Form are **9** pages long. Please make sure you have all the pages.

Your Consent

You are invited to take part voluntarily in this research project.

This participant information document contains detailed information about the research project which has been explained to you verbally. Its purpose is to explain to you as openly and clearly as possible all the procedures involved in this project before you decide whether or not to take part.

Please read this *Participant Information Form* carefully. Feel free to ask questions about any information in the document. You may also wish to discuss the project with a relative or friend or your local health worker. Feel free to do this.

Once you understand what the project is about and if you agree to take part in it, you will be asked to sign the Consent Form. By signing the Consent Form, you indicate that you understand the information and that you give your consent to participate in the research project.

You will be given a copy of the Participant Information and Consent Form to keep as a record.

What is the study about?

The aim of this project is to have an impact on the eventual prevention and treatment of lifestyle diseases in Africans from South Africa. New knowledge regarding the relationship between higher nervous system activity implicating cardiovascular, metabolic and psychological well-being will improve understanding and change strategies at the roots of treatment and prevention of lifestyle diseases.

Our research has shown that lifestyle diseases in urban Africans present higher obesity levels, high blood pressure or hypertension prevalence rates and the experiencing of more stress. This pattern is enhanced during psychosocial stress/urbanisation in participants with a specific coping style. Hence, the planned SABPA project, which is the first study in South Africa where coping and direct markers of in Africans will be measured.

Purpose of study

The purpose of this study is to repeat our previous measurements to investigate biological markers associated with higher nervous system activity in urban teachers with a specific coping style.

To investigate the relationship between blood pressure, inflammation, obesity, stress and coping in more detail we are going to perform this study in 409 men and women from the North-West province, aged 25-65 years. A comprehensive assessment of the cardiovascular and nervous systems by means of non-invasive painless techniques will be performed and blood and saliva samples will be taken by an experienced research doctor and nurse to determine your blood sugar, cardiovascular, inflammation and stress hormone levels amongst other health markers.

Procedures

All measurements are performed in the Metabolic Unit (lipid clinic) of the University. A researcher has explained the entire procedure in detail and while you are reading this information document you have time to ask questions and to have clarified matters. If you are fine with the explained procedure you are requested to sign a *consent form (at the end of this document). Remember all personal data will be handled with care and remain confidential.

**By consenting to participate in this study, you consent to the storage and later analysis and testing of your stored blood samples for the purposes noted above. Your blood will also be tested for preliminary results on HIV status, since your HIV status may directly influence the main purposes of this study. If you would like to know what your HIV-status is, we will provide it. If tested positive we will refer you to your doctor and he/she will perform the necessary tests which will allow you to apply for chronic medication benefit. Also, the blood cells from your donated blood sample will be used to investigate the molecular genetics of higher nervous system activity and type 2 diabetes in order to enable pre-symptomatic diagnosis of hypertension and diabetes in the long term.*

Why was I chosen? Educators are exposed to changing curricula and disciplinary problems whilst living in an urbanised environment adding to higher stress experiencing and nervous system activity.

How was I chosen?

Inclusion criteria*:*
Phase I: black African educators aged 25-65 years (male and female)
Phase II: white African educators aged 25-65 years (male and female).

Exclusion criteria: *pregnancy, lactation, alpha and beta blockers, temperature >37°C. You can not be included if you have donated blood or been vaccinated in the previous 3 months.*

What will be expected of me?

You, as participant will be screened once by a registered nurse to be eligible complying with the inclusion criteria. The following procedures will be followed:

- Recruitment and informed sessions with all participants will be done two months prior to the study (October - November 2010, Phase I, and November, 2011, Phase II) and informed consent forms will be signed.
- After selection of all participants, the details of the project will be discussed with you in English or your home language, i.e. what the exact objectives of the study are, what procedures will be taken and what will be expected from each of you (e.g. overnight stay, resting blood pressure procedures and fasting urine and blood samples are required, importance of complying with the correct sampling methods, incentives). You will be given the opportunity to ask questions.
- Data collection for each participant will involve two days (15min in the morning and 2½ hours in the evening) on Day I; and 2 hours on Day II):

DAY I

- On day I between 07:00-08:00, the blood pressure apparatus, which will measure your blood pressure and heart function as well as a physical activity apparatus will be applied to your arm and waist at your school and you can then resume your normal daily activities. Urine sampling (24h) and 24h diets will commence.
- At the end of Day I (± 15:30) you will be transported from your schools to overnight in the Metabolic Unit Research Facility of the North-West University. This unit is a research unit for human studies and equipped with 10 well furnished bedrooms, a kitchen, two bathrooms and a television room. Each of you will be subjected to the following procedures:
  - At the end of Day I ± 16:00 you will be welcomed and each of you will receive your own private bedroom and eye measurements will commence.
  - Pre-cunselling for HIV/AIDS will be done. All other apparatus will be shown and the procedures, which will be done, will be explained again and you will receive dinner.
  - After dinner, the psychosocial questionnaires will be completed under supervision of registered clinical psychologists/postgraduate students. Completion of questionnaires will take approximately 40 min, From 22:00 you will be fasting, therefore, this will be you last meal for Day I as you must be fasting on Day II for obtaining good results.
  - Thereafter, you can relax and watch television or socialize with your co-participants. It will be wise to go to bed not later than 22:00 as the blood pressure apparatus will take measurements every hour during the night and it can be tiring.

DAY II

- At 07:00 on Day II in the anthropometric station your weight, height and body circumferences will be measured.
- The blood pressure apparatus will be removed at 07:30 and the last urine sample collected.
- Next the cardiovascular measurements will follow consisting of three separate procedures:
  - Firstly, after being in semi-Fowlers position for 5 minutes your blood pressure will be taken in duplicate with the sphygmomanometer (the same as used at clinics) with a resting period of 5 minutes in between.

- Secondly, our registered research nurse will measure the ECG which measures heart function, with 12 leads, which will be placed into position on your rib cage/front part of the body.
- Thirdly, the assessment of pulse wave velocity will follow, i.e. giving an indication of how stiff your vessel walls are. The stiffer your vessel wall is the faster the blood travels from one point of your body to another. These painless measurements will require two technicians using blunt probes (tonometer) putting light pressure on the neck and on the foot to measure the velocity of the pulse waves. This takes only a few minutes.
- Laslty, an ultrasound device will be taken of your arteries in the neck with a blunt probe to indicate the intrinsic thickness of your arteries which contributes to high blood pressure.

➢ A once-off blood sample of 47,5 will be obtained between 08:30 - 09:00 from a vein in your dominant arm. The two stressors you will be exposed to for one minute include:

You have reached the end of the sampling phase.

➢ Immediate feedback on your HIV/AIDS status, obesity levels, blood pressure and blood glucose/sugar values will be given. *HIV/AIDS post-test counselling will be arranged if you are tested positive.*

➢ **Thank you for your participation! You now will have the opportunity to shower and a take away breakfast will be given**.

➢ You will now be transported back to your school and after one week you will receive your 24-hour blood pressure, 12 lead ECG and eye reports as well as sleeping disturbances/sleep apnea risk.

Possible Risks

The measurements performed in our study will include only non-invasive techniques that are not expected to reveal any risks but might cause little discomfort. The taking of blood samples is an invasive procedure with a minimal risk of bleeding. Thus the procedure may cause only a few seconds of light discomfort. All tests will be performed by experienced research nurses of our department. There may be additional unforeseen or unknown risks.

Precautions to protect the participant

The Metabolic Unit facility of the NWU is fully equipped, and in case of an emergency which could not be handled by the registered nurse, the supervising medical doctor Emile Kotzé will be contacted. Dr. Kotzé was notified before the study commenced that this study will be taking place, and that there is a slight possibility that he may be contacted. Supporting medical treatment care facilities will be at hand anytime if needed.

Other Treatments Whilst on Study

It is important to tell the research staff about any treatments or medications you may be taking, including non-prescription medications, vitamins or herbal remedies during your participation in the study.

Incentives

1. All educators will receive feedback on their health profile and if necessary references will be given to physicians/clinics/hospitals.

2. Blood pressure and ECG monitoring report (normally costing £45.00). Your benefit of participation is a comprehensive assessment of the cardiovascular and metabolic condition including investigation of blood pressure, inflammatory status and psychological well-being. These examinations will help us to assess the degree of vascular impairment of the arteries and to predict your risk of possible cardiovascular events such as heart attacks and stroke. The results may assist your doctor in decision making for further treatment or for instituting preventive measures. Our study will also contribute to the identification of possible factors leading to high blood pressure. As 24 hour ambulatory blood pressure monitoring is required for the diagnosis of hypertension, medical aids insist on this method of diagnosis to qualify for chronic medication. Additional testing could also reveal illnesses of a chronic nature and would serve as a motivation to qualify for chronic medication, such as metabolic syndrome, anti-inflammatory and cholesterol-lowering drugs.
3. Monetary incentive as compensation for discomfort and token of appreciation of R100.00 / US$±14))
4. Diet for 24 hours (R150.00 / Two breakfast (± US$121.50).

Privacy, Confidentiality and Disclosure of Information

By consenting to participate in this study, you consent to the storage and later analysis and testing of your stored blood samples for purposes noted above. Your blood samples will be discarded immediately after analysis. All information provided by you and the results of tests will be treated in the strictest confidence, and will only be used for the purpose of this research project. It will only be disclosed with your permission, except as required by law. The results of your medical tests will be labelled only with a code number, and will be stored separately from any identifying information. When the results are analysed we will be looking for differences between groups of people, not at the results of individuals. No information that could identify any person taking part in the study will be revealed when the results are reported.

Participation is Voluntary

Participation in any research project is voluntary. If you do not wish to take part you are not obliged to. If you decide to take part and later change your mind, you are free to withdraw from the project at any stage.

Your decision whether to take part or not to take part, or to take part and then withdraw, will not affect your routine treatment, your relationship with those treating you or your relationship with the North-West University.

Before you make your decision, a member of the research team will be available so that you can ask any questions you have about the research project. You can ask for any information you want. Sign the Consent Form only after you have had a chance to ask your questions and have received satisfactory answers.

If you decide to withdraw from this project, please notify a member of the research team before you withdraw.

Ethical Guidelines

This project will be carried out according to Ethical Guidelines of the Helsinki declaration from 2004, with additional notes in 2002. This statement has been developed to protect the interests of people who agree to participate in human research studies.

The ethical aspects of this research project have been approved by the Human Research Ethics Committee of ***North-West University Potchefstroom:***
(NEW-EC): 0003603S6

Further Information or Any Problems

If you require further information or if you have any problems concerning this project, you can contact the principal researcher or *the other* researchers responsible for this project.

Prof Leoné Malan (018-299 2438)

Project Leader Signature 

PART 2

To the subject signing the consent as in part 3 of this document

You are invited to participate in a research project as described in paragraph 2 of Part 1 of this document. It is important that you read/listen to and understand the following general principles, which apply to all participants in our research project:

1. **Participation in this project is voluntary.**
2. **It is possible that you personally will not derive any benefit from participation in this project, although the knowledge obtained from the results may be beneficial to other people.**
3. **You will be free to withdraw from the project at any stage without having to explain the reasons for your withdrawal. However, we would like to request that you would rather not withdraw without a thorough consideration of your decision, since it may have an effect on the statistical reliability of the results of the project.**
4. **The nature of the project, possible risk factors, factors which may cause discomfort, the expected benefits to the subjects and the known and the most probable permanent consequences which may follow from your participation in this project, are discussed in Part 1 of this document.**
5. **We encourage you to ask questions at any stage about the project and procedures to the project leader or the personnel, who will readily give more information. They will discuss all procedures with you.**

PART 3

Consent

Title of the project: "**THE SABPA STUDY** (**S**ympathetic activity and **A**mbulatory **B**lood **P**ressure in **A**fricans)".

I, the undersigned…… (full names)
read / listened to the information on the project in PART 1 and PART 2 of this document and I declare that I understand the information. I had the opportunity to discuss aspects of the project with the project leader and I declare that I participate in the project as a volunteer. I hereby give my consent to be a subject in this project.

(Signature of the subject)

Signed at ..... on ……2010/11

Witnesses

1. .....

2. . .....

Signed at ..... on ……2010/11
2007/8

APPENDIX C: CYBRIDS STUDY CONSENT FORM

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

HREC Stamp

PARTICIPANT INFORMATION LEAFLET AND RE-CONSENT FORM FOR ALL PARTICIPATING IN THE CYBRID STUDY

TITLE OF THE RESEARCH PROJECT: *In vitro* functional investigation of the effect of rare mtDNA variants in cytoplasmic hybrid cells

REFERENCE NUMBERS: (Ethics approval number) NWU-00358-16-S1 (current study) and NWU-00102-12-A1, "Biotransformation and Oxidative Stress Assessment" (previous study, 2013).

PRINCIPAL INVESTIGATOR: Prof FH van der Westhuizen (PhD Biochemistry)
STUDY CO-ORDINATOR: Mrs Cecile Cook (National diploma Medical Technology)

ADDRESS: Focus Area for Human Metabolomics, Building F3, North-West University, Potchefstroom Campus

CONTACT NUMBER: 018 299 2305/2318

You have previously taken part in the "Biotransformation and Oxidative Stress Assessment" study in 2013. In that study, subjects who were willing to take part had their oxidative stress status studied along with how well their body was able to naturally detoxify foreign and natural toxins. From the results of that study we were able to gather very valuable information in mostly healthy individuals for future studies. Some interesting information that was obtained from this was the inherited (genetic) information, which had to do with the way that energy can be processed.

We would like to follow up on that previous study and would like to invite you to take part in this new study by allowing us to use your blood sample that was obtained during the previous study.

Please take some time to read the information given here, which will explain the details of this new project. Please ask the researcher any questions about any part of this project that you do not fully understand. It is very important that you clearly understand what this research is all about and how you can be involved. Also, your participation is **entirely voluntary** and you are free to decline to participate. If you say no, this will not have any negative effect on you whatsoever. You are also free to withdraw from the study at any point, even if you agreed to take part at the start.

This study has been approved by the **Health Research Ethics Committee of the Faculty of Health Sciences of the North-West University (NWU ethics approval nr. NWU 00358-16-S1)** and will be carried out according to the ethical guidelines and principles of the international Declaration of Helsinki and the ethical guidelines of the National Health Research Ethics Council. It might be necessary for the research ethics committee members or relevant authorities to have a look at the research records.

8.3.7 What is this research study all about?

In short, this study wants to investigate the effect of some small genetic differences often found on the *mitochondrial* DNA of healthy people.

In our cells there are two inherited genetic parts: the first part is in the nucleus (called *nuclear DNA*), which is the largest and most important genetic feature inherited from both our parents. The second part, which is less common knowledge, is called *mitochondrial DNA (mtDNA)* and is found in the part of our cells called “mitochondria”. Mitochondria are the energy producing elements of our cells, so they are a very important part of our cells. Mitochondrial DNA are quite unique as they are only inherited from our mothers and, compared to nuclear DNA, are very small (about 6/10 000 the size!). They are very important for energy metabolism and for that reason we have been interested in their role in the energy demanding functions such as biotransformation and oxidative stress.

In the past years we have found some differences between healthy people’s mtDNA, that are not known to affect health, but might still cause slight differences in the way we use the energy from what we eat. We want to investigate these differences using a new technique. The technique involves making “cytoplasmic hybrid” (or *cybrid* for short) cells by using mtDNA that can be obtained from a part of blood. In the figure below a bit more detail of how this can be done is shown. We will be happy to explain this process to you, because this might look very hard to understand!

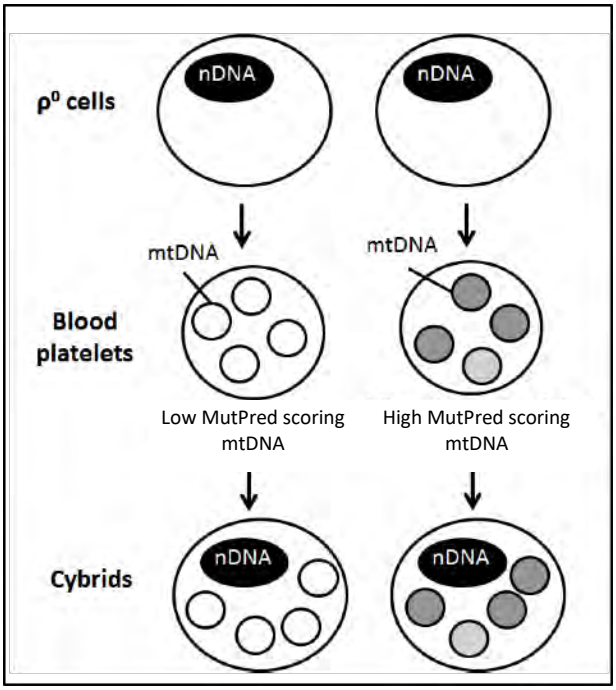


Figure 1. The process of making of cybrid cells from blood platelets using donor (rho 0) cells. The picture indicates an example where different types of mtDNA can be compared to others.

We plan to:

- Use stored blood samples to transfer mtDNA found in one component of the blood (called the blood platelets) into other cells that do not contain mtDNA (called ρ^0 cells) (see Figure 1). This forms a cybrid cell that now can be used to compare the effect of the different types of mtDNA. This is almost like testing different car batteries using the same car, where in the experiment the batteries from different cars are taken out and put into the same model of car – then tests are performed on this one car type with the different types of batteries.
- Do tests to see if these different types of mtDNA have an effect on the way we process energy in our cells.

8.3.8 **Why have you been invited to participate?**

You have been invited to take part because, in the previous study, the mtDNA data that we got from your blood showed that your mtDNA is a type that we would like to compare to other types of mtDNA.

8.3.9 **What will be expected of you?**

If you understand the purpose of the study and what we want to do, and agree that we can use the blood still available from the previous study, give us written consent to do so by signing this from. We will **not** be collecting any new blood samples from you, we will be making use of the previously collected blood samples. We will contact you again after **one week** in order for you to decide whether or not you would like to consent to this.

8.3.10 **Will you gain anything from taking part in this research?**

There will be no direct benefit from this study, but by allowing us to use your blood sample, the indirect benefit is that it will help medical science and genetics better understand if and how small differences between our mtDNA can have an effect on the way we use energy from our diet. This

could also help us to better understand if these effects have an impact on many rare and common diseases.

8.3.11 Are there risks involved in your taking part in this research?

- There are no physical or other risks for you by taking part in this study. We do not require anything other than consent from you.
- No new genetic data will be created and there is no chance that your identity can become known from the methods given here.
- There are only benefits to this study.

How will we protect your confidentiality and who will see your findings?
Your identity is already protected in that a laboratory number has been assigned to your samples during the previous study. We will use only this number and never your name. Only the researchers and the supervisory doctor from the previous study will be able to look at your results. These persons have all signed confidentiality agreements. Your privacy will be respected by allowing only three persons (Prof Francois van der Westhuizen, Mr Lardus Erasmus and Mrs Cecile Cook) to access the personal information that links you to a specific laboratory number. Findings will be kept safe by storing hard copies in locked cupboards in the researcher's office and electronic data will be password protected. The data will be stored for a minimum of seven years but we will store the samples as long as possible, unless you want it destroyed.

What will happen with the data/samples?
The data will be stored for 7 years and the material will be stored only at the Centre for Human Metabolomics, North-West University, Potchefstroom, until you request (in writing) the destruction of this material at any time.

None of the biological samples will be transported or tested abroad, they will only be stored and tested at the Centre for Human Metabolomics, Potchefstroom campus, NWU.

How will you know about the findings?
A final information session will be arranged during which the findings of the study will be reported to you and the other participants. It is expected that this will happen before the end of 2017. A report with a summary of the key findings will be made available electronically as well as in hard copy. Participants that have access to email will receive the electronic copy, others will receive a hard copy during the final information session. During this session the results will be explained and participants will have the opportunity to ask questions.

Will you be paid to take part in this study and are there any costs involved?
There is no remuneration for participating in the follow up study.

Is there anything else that you should know or do?

- You can contact Prof Francois van der Westhuzien at 018 299 2318 or Francois.vanderWesthuizen@nwu.ac.za if you have any further queries or encounter any problems.

- You can also contact Mrs Cecile Cook at 018 299 2305 or Cecile.Cook@nwu.ac.za if you have questions relating to this or the previous study.
- You can contact the Health Research Ethics Committee via Mrs Carolien van Zyl at 018 299 1206; carolien.vanzyl@nwu.ac.za if you have any concerns or complaints that have not been adequately addressed by the researcher.
- You will receive a copy of this information and consent form for your own records.

Declaration by participant

By signing below, I …….. agree to take part in a research study entitled:

***In vitro* functional investigation of the effect of rare mtDNA variants in cytoplasmic hybrid cells**

I declare that:

- I have read this information and consent form and it is written in a language with which I am fluent and comfortable.
- I have had a chance to ask questions and all my questions have been adequately answered.
- I understand that taking part in this study is **voluntary** and I have not been pressurised to take part.
- I may choose to leave the study at any time and will not be penalised or prejudiced in any way.
- I may be asked to leave the study before it has finished, if the researcher feels it is in my best interests, or if I do not follow the study plan, as agreed to.
- I may be contacted in case a discovery is made that could impact on my health:
  Yes ☐ No ☐
- I may be contacted if further participation or re-consent is required when new discoveries have been made: Yes ☐ No ☐

Signed at (place) …….. on (date) …….. 20....

…… ……

Signature of participant **Signature of witness**

Declaration by investigator

I (name) …… declare that:

- I explained the information in this document to …….
- I encouraged him/her to ask questions and took adequate time to answer them.
- I am satisfied that he/she adequately understands all aspects of the research, as discussed above
- I did/did not use an interpreter.

Signed at (place) ..... on (date) …….. 20....

…… ……

Signature of investigator **Signature of witness**

Declaration by independent person

I (name) …….. declare that:

- I explained the information in this document to …….
- I encouraged him/her to ask questions and took adequate time to answer them.
- I am satisfied that he/she adequately understands all aspects of the research, as discussed above
- I did/did not use an interpreter.

Signed at (place) ..... on (date) …….. 20....

…… ……

Signature of independent person **Signature of witness**

APPENDIX D: SUPPLEMENTARY MATERIALS FOR ARTICLE PUBLISHED IN JOURNAL OF GENETICS AND GENOMICS (CHAPTER 5)

Supplimentary Table 1: Important phenotypical and genetic data for each participant

Gender	24 hr Systolic / Diastolic ABPM (mmHg)	HbA1c (%)	Blood pressure and blood glucose classifications	Complete cohort			Variants with a MutPred pathogenicity score above 0.3			Variants with a MutPred pathogenicity score above 0.5			Variants with a MutPred pathogenicity score above 0.6		
				Mut. load	Num of var	Adj load	Mut. load	Num of var	Adj load	Mut. load	Num of var	Adj load	Mut. load	Num of var	Adj load
Haplogroup L participants															
Male	136 / 82	5.5	Hyper and optimal	4.97	13	0.38	4.24	9	0.47	1.92	3	0.64	1.92	3	0.64
Male	148 / 83	5.7	Hyper and pre-diabetic	7.63	19	0.40	6.93	15	0.46	3.38	6	0.56	1.82	3	0.61
Male	152 / 97	6.4	Hyper and pre-diabetic	4.65	14	0.33	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	124 / 74	5.7	Hyper and pre-diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	145 / 89	6.1	Hyper and pre-diabetic	2.67	8	0.33	2.19	5	0.44	0.56	1	0.56	0.00	0	0.00
Female	143 / 75	5.8	Hyper and pre-diabetic	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	181 / 112	6.1	Hyper and pre-diabetic	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Female	136 / 82	5.3	Hyper and optimal	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	124 / 70	6.4	Hyper and pre-diabetic	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Female	121 / 78	6	Hyper and pre-diabetic	4.87	14	0.35	4.21	10	0.42	0.61	1	0.61	0.61	1	0.61
Female	117 / 77	5	Hyper and optimal	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	145 / 93	6.1	Hyper and pre-diabetic	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Male	158 / 101	12.3	Hyper and diabetic	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Male	143 / 90	5.9	Hyper and pre-diabetic	3.21	10	0.32	2.29	5	0.46	1.12	2	0.56	0.61	1	0.61
Male	147 / 97	5.6	Hyper and optimal	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Male	146 / 95	10.4	Hyper and diabetic	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Male	137 / 78	5	Hyper and optimal	4.42	13	0.34	3.69	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	157 / 105	6.2	Hyper and pre-diabetic	4.65	14	0.33	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	124 / 81	6.3	Hyper and pre-diabetic	4.89	14	0.35	4.23	10	0.42	1.12	2	0.56	0.61	1	0.61
Male	144 / 91	6.5	Hyper and diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	147 / 75	5.4	Hyper and optimal	5.12	13	0.39	4.40	9	0.49	2.07	3	0.69	2.07	3	0.69
Male	138 / 90	5.5	Hyper and optimal	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Male	131 / 86	8.8	Hyper and diabetic	8.38	20	0.42	7.68	16	0.48	4.00	7	0.57	2.44	4	0.61
Male	142 / 100	5.8	Hyper and pre-diabetic	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	113 / 71	5.3	Normal and optimal	5.54	14	0.40	5.06	11	0.46	1.87	3	0.62	1.87	3	0.62

Male	154 / 103	6.8	Hyper and diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	121 / 79	5.6	Normal and optimal	7.11	18	0.40	6.39	14	0.46	2.97	5	0.59	1.91	3	0.64
Female	151 / 78	5.7	Hyper and pre-diabetic	1.85	7	0.26	1.16	3	0.39	0.00	0	0.00	0.00	0	0.00
Female	110 / 64	5	Normal and optimal	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Female	123 / 73	5.8	Hyper and pre-diabetic	5.58	16	0.35	4.61	11	0.42	0.61	1	0.61	0.61	1	0.61
Female	123 / 74	7.3	Hyper and diabetic	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Female	136 / 85	6.9	Hyper and diabetic	8.41	20	0.42	7.70	16	0.48	4.28	7	0.61	3.21	5	0.64
Female	137 / 75	5.7	Hyper and pre-diabetic	2.64	8	0.33	2.16	5	0.43	0.00	0	0.00	0.00	0	0.00
Female	125 / 82	11.2	Hyper and diabetic	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	133 / 82	5.1	Hyper and optimal	7.68	19	0.40	6.96	15	0.46	3.54	6	0.59	2.47	4	0.62
Female	124 / 79	5.7	Hyper and pre-diabetic	4.42	13	0.34	3.69	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	111 / 65	5.5	Normal and optimal	5.55	14	0.40	4.83	10	0.48	2.07	3	0.69	2.07	3	0.69
Female	132 / 84	6.3	Hyper and pre-diabetic	4.94	13	0.38	4.21	9	0.47	1.46	2	0.73	1.46	2	0.73
Male	136 / 83	6.3	Hyper and pre-diabetic	5.38	15	0.36	4.48	10	0.45	1.85	3	0.62	1.34	2	0.67
Male	112 / 73	5.7	Normal and pre-diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	139 / 90	5.7	Hyper and pre-diabetic	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	152 / 108	6.8	Hyper and diabetic	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	145 / 87	6.5	Hyper and diabetic	3.28	9	0.36	2.80	6	0.47	1.15	2	0.57	1.15	2	0.57
Male	131 / 84	5.7	Hyper and pre-diabetic	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Female	154 / 95	8.7	Hyper and diabetic	7.68	19	0.40	6.96	15	0.46	3.54	6	0.59	2.47	4	0.62
Female	132 / 86	5.1	Hyper and optimal	3.92	10	0.39	3.44	7	0.49	1.79	3	0.60	1.79	3	0.60
Male	147 / 102	10.1	Hyper and diabetic	4.36	13	0.34	3.63	9	0.40	0.61	1	0.61	0.61	1	0.61
Female	133 / 84	5.7	Hyper and pre-diabetic	6.41	15	0.43	5.69	11	0.52	2.91	4	0.73	2.91	4	0.73
Male	151 / 94	6.6	Hyper and diabetic	4.42	13	0.34	3.69	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	150 / 82	6.8	Hyper and diabetic	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Female	114 / 67	5.6	Normal and optimal	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	154 / 100	7.5	Hyper and diabetic	4.15	12	0.35	3.42	8	0.43	0.61	1	0.61	0.61	1	0.61
Male	130 / 84	5.5	Hyper and optimal	5.20	13	0.40	4.72	10	0.47	2.38	4	0.60	1.29	2	0.64
Male	135 / 87	5.6	Hyper and optimal	5.35	15	0.36	4.67	11	0.42	1.29	2	0.65	1.29	2	0.65
Male	157 / 91	6.3	Hyper and pre-diabetic	5.20	13	0.40	4.72	10	0.47	2.38	4	0.60	1.29	2	0.64
Male	161 / 111	5.5	Hyper and optimal	5.08	14	0.36	4.42	10	0.44	1.31	2	0.66	1.31	2	0.66
Male	138 / 87	6.4	Hyper and pre-diabetic	2.65	8	0.33	2.17	5	0.43	0.00	0	0.00	0.00	0	0.00
Male	111 / 73	5.6	Normal and optimal	3.25	10	0.32	2.50	6	0.42	0.62	1	0.62	0.62	1	0.62
Male	142 / 96	5.4	Hyper and optimal	4.60	14	0.33	3.44	8	0.43	1.12	2	0.56	0.62	1	0.62
Female	126 / 83	5.7	Hyper and pre-diabetic	4.11	11	0.37	3.64	8	0.45	1.20	2	0.60	0.67	1	0.67
Female	144 / 84	5.7	Hyper and pre-diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	133 / 72	6.2	Hyper and pre-diabetic	1.55	6	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	107 / 62	7.6	Normal and diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	123 / 75	5.9	Normal and pre-diabetic	4.40	13	0.34	3.43	8	0.43	1.64	3	0.55	0.58	1	0.58
Male	118 / 79	5.7	Normal and pre-diabetic	5.38	15	0.36	4.48	10	0.45	1.85	3	0.62	1.34	2	0.67

Male	140 / 84	6.1	Hyper and pre-diabetic	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Female	137 / 78	6.2	Hyper and pre-diabetic	6.98	18	0.39	6.26	14	0.45	3.28	6	0.55	1.15	2	0.58
Female	115 / 73	5.3	Normal and optimal	5.00	14	0.36	4.35	10	0.43	1.24	2	0.62	1.24	2	0.62
Male	134 / 96	5.9	Hyper and pre-diabetic	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Male	129 / 83	5.7	Hyper and pre-diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	119 / 79	5.4	Hyper and optimal	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Female	107 / 71	6.2	Normal and pre-diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	126 / 90	5.5	Hyper and optimal	7.26	18	0.40	6.56	14	0.47	3.38	6	0.56	1.82	3	0.61
Male	122 / 75	6	Normal and pre-diabetic	5.00	15	0.33	4.11	10	0.41	1.14	2	0.57	0.63	1	0.63
Male	135 / 86	5.9	Hyper and pre-diabetic	3.94	12	0.33	3.28	8	0.41	0.61	1	0.61	0.61	1	0.61
Female	140 / 87	6	Hyper and pre-diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	119 / 78	5.7	Hyper and pre-diabetic	3.23	9	0.36	2.75	6	0.46	1.21	2	0.61	1.21	2	0.61
Female	116 / 73	5.6	Normal and optimal	2.60	8	0.33	2.13	5	0.43	0.00	0	0.00	0.00	0	0.00
Male	132 / 82	5.8	Hyper and pre-diabetic	5.20	14	0.37	4.72	11	0.43	1.67	3	0.56	1.15	2	0.58
Male	124 / 79	6.3	Normal and pre-diabetic	4.90	13	0.38	4.42	10	0.44	1.31	2	0.66	1.31	2	0.66
Female	117 / 75	5.7	Hyper and pre-diabetic	1.55	6	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	121 / 73	5.4	Hyper and optimal	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Female	144 / 90	5.5	Hyper and optimal	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Female	131 / 71	5.7	Hyper and pre-diabetic	2.11	7	0.30	1.64	4	0.41	0.00	0	0.00	0.00	0	0.00
Female	120 / 70	6.3	Hyper and pre-diabetic	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Female	113 / 73	5.4	Normal and optimal	5.34	15	0.36	4.61	11	0.42	0.61	1	0.61	0.61	1	0.61
Female	127 / 79	5.8	Hyper and pre-diabetic	4.39	14	0.31	3.24	8	0.41	0.51	1	0.51	0.00	0	0.00
Female	130 / 81	5	Hyper and optimal	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Female	123 / 80	5.2	Hyper and optimal	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Female	119 / 74	5.8	Normal and pre-diabetic	2.26	7	0.32	1.79	4	0.45	0.56	1	0.56	0.00	0	0.00
Female	105 / 68	6.1	Normal and pre-diabetic	1.82	6	0.30	1.34	3	0.45	0.52	1	0.52	0.00	0	0.00
Female	116 / 77	5.4	Hyper and optimal	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Female	113 / 74	6.4	Normal and pre-diabetic	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Female	139 / 91	5.7	Hyper and pre-diabetic	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Female	163 / 100	6	Hyper and pre-diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	132 / 85	5.4	Hyper and optimal	1.13	4	0.28	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	162 / 96	6.3	Hyper and pre-diabetic	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Male	148 / 94	9.5	Hyper and diabetic	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Male	128 / 88	6	Hyper and pre-diabetic	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	135 / 94	6.3	Hyper and pre-diabetic	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Male	152 / 96	6.1	Hyper and pre-diabetic	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Male	133 / 88	6.4	Hyper and pre-diabetic	4.52	14	0.32	3.36	8	0.42	1.04	2	0.52	0.00	0	0.00
Male	118 / 78	5.1	Normal and optimal	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	130 / 82	6.4	Hyper and pre-diabetic	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Male	140 / 88	5.6	Hyper and optimal	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00

Male	148 / 92	6.2	Hyper and pre-diabetic	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Male	125 / 85	7.4	Hyper and diabetic	4.90	13	0.38	4.42	10	0.44	1.31	2	0.66	1.31	2	0.66
Female	133 / 83	5	Hyper and optimal	5.12	13	0.39	4.40	9	0.49	2.07	3	0.69	2.07	3	0.69
Female	130 / 84	6.3	Hyper and pre-diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	132 / 90	6.3	Hyper and pre-diabetic	4.42	13	0.34	3.69	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	132 / 85	5.3	Hyper and optimal	3.92	10	0.39	3.44	7	0.49	1.79	3	0.60	1.79	3	0.60
Female	115 / 67	5.2	Normal and optimal	2.84	8	0.35	2.36	5	0.47	0.66	1	0.66	0.66	1	0.66
Male	144 / 92	5.9	Hyper and pre-diabetic	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	145 / 85	6.2	Hyper and pre-diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	179 / 88	5.6	Hyper and optimal	2.97	10	0.30	2.05	5	0.41	0.51	1	0.51	0.00	0	0.00
Male	199 / 126	5.9	Hyper and pre-diabetic	3.94	12	0.33	3.28	8	0.41	0.61	1	0.61	0.61	1	0.61
Female	130 / 77	5.5	Hyper and optimal	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Female	129 / 78	5.3	Hyper and optimal	1.55	6	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	127 / 81	5.9	Hyper and pre-diabetic	4.32	13	0.33	3.17	7	0.45	1.29	2	0.64	0.78	1	0.78
Female	115 / 64	6	Normal and pre-diabetic	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Female	129 / 82	6.1	Hyper and pre-diabetic	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Male	140 / 84	5.5	Hyper and optimal	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Male	131 / 82	5.7	Hyper and pre-diabetic	6.83	17	0.40	6.13	13	0.47	3.38	6	0.56	1.82	3	0.61
Female	117 / 78	5.3	Hyper and optimal	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	110 / 72	5.3	Normal and optimal	4.03	10	0.40	3.72	8	0.46	1.84	3	0.61	1.34	2	0.67
Female	111 / 72	5.2	Normal and optimal	2.65	8	0.33	2.17	5	0.43	0.00	0	0.00	0.00	0	0.00
Female	111 / 73	5.5	Normal and optimal	5.75	15	0.38	5.28	12	0.44	1.79	3	0.60	1.29	2	0.64
Female	130 / 84	6.2	Hyper and pre-diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	122 / 72	6.1	Normal and pre-diabetic	3.23	9	0.36	2.75	6	0.46	1.21	2	0.61	1.21	2	0.61
Male	172 / 111	5.4	Hyper and optimal	3.23	9	0.36	2.75	6	0.46	1.21	2	0.61	1.21	2	0.61
Male	112 / 72	5.8	Normal and pre-diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	168 / 90	5.2	Hyper and optimal	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	139 / 92	5.2	Hyper and optimal	4.42	13	0.34	3.69	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	145 / 91	5.1	Hyper and optimal	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Female	116 / 78	5.3	Hyper and optimal	4.71	13	0.36	4.24	10	0.42	1.12	2	0.56	0.61	1	0.61
Female	115 / 75		Hyper and	7.11	18	0.40	6.39	14	0.46	2.97	5	0.59	1.91	3	0.64
Male	114 / 77	5.7	Normal and pre-diabetic	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Male	154 / 94	6.6	Hyper and diabetic	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Male	140 / 94	5.4	Hyper and optimal	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Male	117 / 72	5.4	Normal and optimal	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Female	148 / 93	7.4	Hyper and diabetic	6.04	17	0.36	4.98	11	0.45	1.78	3	0.59	1.26	2	0.63
Female	130 / 82	6.4	Hyper and pre-diabetic	4.85	14	0.35	3.97	9	0.44	1.64	3	0.55	0.61	1	0.61
Female	130 / 85	5.6	Hyper and optimal	4.42	13	0.34	3.69	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	117 / 80	5.3	Hyper and optimal	6.79	16	0.42	6.48	14	0.46	2.86	5	0.57	1.82	3	0.61
Female	129 / 86	5.5	Hyper and optimal	3.60	11	0.33	2.90	7	0.41	0.55	1	0.55	0.00	0	0.00

Female	135 / 85	5.7	Hyper and pre-diabetic	3.73	10	0.37	3.25	7	0.46	1.71	3	0.57	1.21	2	0.61
Female	145 / 88	6.1	Hyper and pre-diabetic	7.11	18	0.40	6.39	14	0.46	2.97	5	0.59	1.91	3	0.64
Female	129 / 70	5.7	Hyper and pre-diabetic	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	137 / 76	5.5	Hyper and optimal	4.42	13	0.34	3.69	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	141 / 94	5.4	Hyper and optimal	4.81	13	0.37	4.34	10	0.43	1.23	2	0.61	1.23	2	0.61
Male	140 / 89	5.9	Hyper and pre-diabetic	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Male	117 / 77	6.2	Normal and pre-diabetic	1.98	6	0.33	1.50	3	0.50	0.68	1	0.68	0.68	1	0.68
Male	128 / 76	6.1	Normal and pre-diabetic	3.97	12	0.33	3.25	8	0.41	0.61	1	0.61	0.61	1	0.61
Male	122 / 82	5.1	Hyper and optimal	3.21	10	0.32	2.29	5	0.46	1.12	2	0.56	0.61	1	0.61
Male	146 / 100	6.1	Hyper and pre-diabetic	4.42	13	0.34	3.69	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	124 / 80	5.2	Hyper and optimal	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Male	135 / 87	5.5	Hyper and optimal	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Male	123 / 79	5.9	Normal and pre-diabetic	5.38	15	0.36	4.48	10	0.45	1.85	3	0.62	1.34	2	0.67
Male	158 / 101	9.6	Hyper and diabetic	3.92	10	0.39	3.44	7	0.49	1.79	3	0.60	1.79	3	0.60
Male	125 / 82	9.4	Hyper and diabetic	4.42	13	0.34	3.69	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	114 / 68	5.6	Normal and optimal	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Male	124 / 79	6	Normal and pre-diabetic	5.86	16	0.37	4.97	11	0.45	2.33	4	0.58	1.83	3	0.61
Female	160 / 98	5.7	Hyper and pre-diabetic	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	181 / 112	9.6	Hyper and diabetic	4.42	13	0.34	3.69	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	121 / 66	5.3	Hyper and optimal	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	111 / 65	6.2	Normal and pre-diabetic	5.16	14	0.37	4.68	11	0.43	1.11	2	0.56	0.61	1	0.61
Female	113 / 77	5.3	Hyper and optimal	4.40	13	0.34	3.43	8	0.43	1.64	3	0.55	0.58	1	0.58
Male	161 / 97	5.9	Hyper and pre-diabetic	4.21	12	0.35	3.48	8	0.44	1.16	2	0.58	0.61	1	0.61
Female	122 / 77	5.9	Hyper and pre-diabetic	4.93	14	0.35	4.20	10	0.42	0.61	1	0.61	0.61	1	0.61
Male	124 / 79	6	Normal and pre-diabetic	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Female	104 / 63	5.2	Normal and optimal	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	160 / 84	12.8	Hyper and diabetic	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Female	131 / 82	6.5	Hyper and diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	112 / 74	5.6	Normal and optimal	4.66	13	0.36	4.18	10	0.42	0.61	1	0.61	0.61	1	0.61
Male	161 / 108	5.8	Hyper and pre-diabetic	5.23	14	0.37	4.24	9	0.47	1.92	3	0.64	1.92	3	0.64
Female	126 / 82	5	Hyper and optimal	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Female	120 / 74	5.2	Hyper and optimal	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Male	158 / 94	5.8	Hyper and pre-diabetic	4.42	13	0.34	3.69	9	0.41	0.61	1	0.61	0.61	1	0.61
Male	138 / 93	6.7	Hyper and diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	129 / 82	5.6	Hyper and optimal	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	124 / 83	6.1	Hyper and pre-diabetic	5.34	15	0.36	4.61	11	0.42	0.61	1	0.61	0.61	1	0.61
Female	126 / 81	6.2	Hyper and pre-diabetic	3.46	10	0.35	2.89	7	0.41	0.51	1	0.51	0.00	0	0.00
Female	122 / 68	5.5	Hyper and optimal	7.26	18	0.40	6.56	14	0.47	3.38	6	0.56	1.82	3	0.61
Female	131 / 86	4.8	Hyper and optimal	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Female	145 / 78	5.8	Hyper and pre-diabetic	5.34	15	0.36	4.61	11	0.42	0.61	1	0.61	0.61	1	0.61

Male	131 / 87	5.7	Hyper and pre-diabetic	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Male	111 / 72	5.5	Normal and optimal	4.93	14	0.35	4.20	10	0.42	0.61	1	0.61	0.61	1	0.61
Male	122 / 74	5.2	Normal and optimal	4.66	14	0.33	3.77	9	0.42	1.14	2	0.57	0.63	1	0.63
Male	132 / 78	9.2	Hyper and diabetic	4.38	13	0.34	3.72	9	0.41	0.61	1	0.61	0.61	1	0.61
Female	121 / 74	5.5	Hyper and optimal	1.55	6	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	127 / 73	5.2	Hyper and optimal	4.93	14	0.35	4.20	10	0.42	0.61	1	0.61	0.61	1	0.61
Male	153 / 90	5.6	Hyper and optimal	4.56	13	0.35	4.09	10	0.41	0.61	1	0.61	0.61	1	0.61
Male	135 / 98	5.6	Hyper and optimal	5.54	14	0.40	5.06	11	0.46	1.87	3	0.62	1.87	3	0.62
Male	163 / 97	6.3	Hyper and pre-diabetic	2.18	7	0.31	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Haplogroup MN participants															
Male	133 / 93	5.5	Hyper and optimal	1.35	5	0.27	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	123 / 68	5.6	Hyper and optimal	3.43	8	0.43	3.02	6	0.50	1.81	3	0.60	1.26	2	0.63
Female	190 / 119	5.2	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	128 / 85	5.6	Hyper and optimal	2.09	5	0.42	1.93	4	0.48	1.11	2	0.55	0.00	0	0.00
Female	131 / 72	5.3	Hyper and optimal	4.11	11	0.37	3.33	7	0.48	1.71	3	0.57	1.20	2	0.60
Female	123 / 83	5.7	Hyper and pre-diabetic	0.99	3	0.33	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	120 / 70	5	Hyper and optimal	2.06	6	0.34	1.65	4	0.41	0.00	0	0.00	0.00	0	0.00
Female	111 / 67	5.5	Normal and optimal	3.13	8	0.39	2.86	6	0.48	1.22	2	0.61	1.22	2	0.61
Female	110 / 70	5.4	Normal and optimal	3.88	9	0.43	3.61	7	0.52	1.97	3	0.66	1.97	3	0.66
Male	124 / 82	6.1	Hyper and pre-diabetic	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	132 / 83	5.6	Hyper and optimal	1.13	3	0.38	1.13	3	0.38	0.00	0	0.00	0.00	0	0.00
Male	126 / 81	5.7	Hyper and pre-diabetic	2.53	7	0.36	2.11	5	0.42	0.00	0	0.00	0.00	0	0.00
Female	113 / 69	6.3	Normal and pre-diabetic	4.11	11	0.37	3.33	7	0.48	1.71	3	0.57	1.20	2	0.60
Female	113 / 70	5.7	Normal and pre-diabetic	2.26	6	0.38	1.92	4	0.48	0.61	1	0.61	0.61	1	0.61
Female	111 / 69	5.3	Normal and optimal	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	118 / 72	5.4	Normal and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	112 / 69	5.7	Normal and pre-diabetic	0.99	3	0.33	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	113 / 72	5.6	Normal and optimal	0.99	3	0.33	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	116 / 76	5.3	Hyper and optimal	4.94	12	0.41	4.31	9	0.48	1.27	2	0.64	0.77	1	0.77
Female	117 / 71	5.1	Normal and optimal	2.78	8	0.35	2.16	5	0.43	0.51	1	0.51	0.00	0	0.00
Female	120 / 73	5.1	Hyper and optimal	0.99	3	0.33	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	130 / 79	5.4	Hyper and optimal	1.44	4	0.36	1.17	3	0.39	0.00	0	0.00	0.00	0	0.00
Female	120 / 74	5.5	Hyper and optimal	1.14	4	0.28	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	133 / 81	5.3	Hyper and optimal	3.25	9	0.36	2.82	6	0.47	1.22	2	0.61	1.22	2	0.61
Male	135 / 89	5.2	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	129 / 82	5.3	Hyper and optimal	2.84	7	0.41	2.51	5	0.50	1.19	2	0.60	1.19	2	0.60
Male	158 / 94	6.3	Hyper and pre-diabetic	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	133 / 79	7.8	Hyper and diabetic	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	120 / 79	5.2	Hyper and optimal	2.57	7	0.37	2.06	4	0.52	1.24	2	0.62	1.24	2	0.62
Female	110 / 61	5.2	Normal and optimal	1.50	3	0.50	1.50	3	0.50	0.68	1	0.68	0.68	1	0.68

Male	128 / 69	5.1	Normal and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	168 / 84	5.7	Hyper and pre-diabetic	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	126 / 82	5.7	Hyper and pre-diabetic	2.82	7	0.40	2.56	5	0.51	1.74	3	0.58	1.24	2	0.62
Female	124 / 70	5.3	Hyper and optimal	1.22	4	0.30	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	124 / 79	5.2	Normal and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	117 / 71	5.4	Normal and optimal	1.85	6	0.31	1.22	3	0.41	0.00	0	0.00	0.00	0	0.00
Female	120 / 78	5.2	Hyper and optimal	2.80	7	0.40	2.09	4	0.52	1.27	2	0.64	1.27	2	0.64
Female	133 / 80	5.6	Hyper and optimal	3.99	10	0.40	3.55	7	0.51	1.91	3	0.64	1.40	2	0.70
Male	125 / 78	5.7	Normal and pre-diabetic	1.76	5	0.35	1.43	3	0.48	0.61	1	0.61	0.61	1	0.61
Male	123 / 80	5.8	Hyper and pre-diabetic	2.96	9	0.33	2.24	5	0.45	1.09	2	0.55	0.57	1	0.57
Female	129 / 81	5.4	Hyper and optimal	4.11	11	0.37	3.33	7	0.48	1.71	3	0.57	1.20	2	0.60
Female	112 / 72	5.1	Normal and optimal	5.59	15	0.37	4.87	11	0.44	1.29	2	0.64	1.29	2	0.64
Female	124 / 77	5.6	Hyper and optimal	2.57	7	0.37	2.06	4	0.52	1.24	2	0.62	1.24	2	0.62
Female	125 / 83	5.1	Hyper and optimal	7.12	17	0.42	6.10	12	0.51	2.86	4	0.72	2.86	4	0.72
Female	126 / 74	5.4	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	109 / 70	5	Normal and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	122 / 74	5	Hyper and optimal	2.88	8	0.36	2.45	5	0.49	1.22	2	0.61	1.22	2	0.61
Female	122 / 79	5.6	Hyper and optimal	4.11	11	0.37	3.33	7	0.48	1.71	3	0.57	1.20	2	0.60
Male	118 / 74	5.2	Normal and optimal	2.28	5	0.46	2.12	4	0.53	1.29	2	0.65	1.29	2	0.65
Male	114 / 66	5.1	Normal and optimal	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	124 / 72	5.8	Normal and pre-diabetic	3.88	9	0.43	3.61	7	0.52	1.97	3	0.66	1.97	3	0.66
Male	150 / 91	5.6	Hyper and optimal	1.76	5	0.35	1.43	3	0.48	0.61	1	0.61	0.61	1	0.61
Male	136 / 78	5.2	Hyper and optimal	6.40	15	0.43	5.63	11	0.51	2.86	4	0.72	2.86	4	0.72
Female	123 / 74	6.3	Hyper and pre-diabetic	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	104 / 66	5.6	Normal and optimal	3.52	8	0.44	3.26	6	0.54	2.43	4	0.61	1.88	3	0.63
Female	107 / 61	5.1	Normal and optimal	2.57	7	0.37	2.06	4	0.52	1.24	2	0.62	1.24	2	0.62
Female	137 / 76	5.1	Hyper and optimal	1.36	3	0.45	1.36	3	0.45	0.54	1	0.54	0.00	0	0.00
Male	124 / 81	5.3	Hyper and optimal	2.32	6	0.39	2.06	4	0.52	1.24	2	0.62	1.24	2	0.62
Male	124 / 86	5	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	139 / 87	6.4	Hyper and pre-diabetic	1.76	5	0.35	1.43	3	0.48	0.61	1	0.61	0.61	1	0.61
Female	110 / 66	5.4	Normal and optimal	2.28	5	0.46	2.12	4	0.53	1.29	2	0.65	1.29	2	0.65
Female	102 / 61	5.1	Normal and optimal	1.71	4	0.43	1.71	4	0.43	0.00	0	0.00	0.00	0	0.00
Female	106 / 70	5.8	Normal and pre-diabetic	2.26	6	0.38	1.92	4	0.48	0.61	1	0.61	0.61	1	0.61
Female	122 / 70	5.5	Hyper and optimal	2.06	4	0.51	2.06	4	0.51	1.24	2	0.62	1.24	2	0.62
Female	126 / 73	5.3	Hyper and optimal	2.32	6	0.39	2.06	4	0.52	1.24	2	0.62	1.24	2	0.62
Male	140 / 82	5.2	Hyper and optimal	1.79	5	0.36	1.52	4	0.38	0.00	0	0.00	0.00	0	0.00
Female	121 / 72	5.6	Hyper and optimal	1.80	4	0.45	1.80	4	0.45	0.62	1	0.62	0.62	1	0.62
Female	111 / 66	4.8	Normal and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	135 / 80	5.3	Hyper and optimal	2.62	6	0.44	2.46	5	0.49	1.29	2	0.65	1.29	2	0.65
Female	115 / 77	5	Hyper and optimal	1.59	4	0.40	1.59	4	0.40	0.00	0	0.00	0.00	0	0.00

Female	119 / 71	5.3	Normal and optimal	2.93	7	0.42	2.59	5	0.52	1.28	2	0.64	1.28	2	0.64
Male	127 / 71	5.6	Normal and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	147 / 93	6.1	Hyper and pre-diabetic	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	113 / 75	5.5	Normal and optimal	3.99	10	0.40	3.55	7	0.51	1.91	3	0.64	1.40	2	0.70
Male	122 / 82	5.5	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	126 / 67	5.7	Normal and pre-diabetic	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	115 / 63	5.2	Normal and optimal	1.68	3	0.56	1.68	3	0.56	0.86	1	0.86	0.86	1	0.86
Female	112 / 59	5.2	Normal and optimal	3.58	9	0.40	3.14	6	0.52	1.91	3	0.64	1.40	2	0.70
Female	137 / 86	6.4	Hyper and pre-diabetic	1.24	4	0.31	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	125 / 74	5.3	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	115 / 78	5.6	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	125 / 76	5.8	Normal and pre-diabetic	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	148 / 75	5.7	Hyper and pre-diabetic	2.28	5	0.46	2.12	4	0.53	1.29	2	0.65	1.29	2	0.65
Female	124 / 86	5.2	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	120 / 69	5.2	Normal and optimal	1.13	3	0.38	1.13	3	0.38	0.00	0	0.00	0.00	0	0.00
Female	117 / 70	4.8	Normal and optimal	4.54	14	0.32	3.38	8	0.42	1.06	2	0.53	0.00	0	0.00
Female	127 / 81	5.3	Hyper and optimal	1.60	4	0.40	1.43	3	0.48	0.61	1	0.61	0.61	1	0.61
Female	108 / 66	5	Normal and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	120 / 72	5.4	Normal and optimal	1.67	4	0.42	1.67	4	0.42	0.51	1	0.51	0.00	0	0.00
Female	123 / 81	5.6	Hyper and optimal	3.58	9	0.40	3.14	6	0.52	1.91	3	0.64	1.40	2	0.70
Female	140 / 81	5.8	Hyper and pre-diabetic	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	117 / 74	5.1	Normal and optimal	2.93	7	0.42	2.59	5	0.52	1.28	2	0.64	1.28	2	0.64
Female	115 / 71	5.4	Normal and optimal	0.99	3	0.33	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	117 / 66	6.1	Normal and pre-diabetic	0.99	3	0.33	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	113 / 70	5.4	Normal and optimal	1.37	4	0.34	1.21	3	0.40	0.00	0	0.00	0.00	0	0.00
Male	132 / 82	5.6	Hyper and optimal	3.43	8	0.43	3.01	6	0.50	1.36	2	0.68	1.36	2	0.68
Female	105 / 73	5.2	Normal and optimal	0.99	3	0.33	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	138 / 85	5.1	Hyper and optimal	4.11	11	0.37	3.33	7	0.48	1.71	3	0.57	1.20	2	0.60
Female	106 / 72	5.1	Normal and optimal	1.49	3	0.50	1.49	3	0.50	0.67	1	0.67	0.67	1	0.67
Male	135 / 90	5.5	Hyper and optimal	1.76	5	0.35	1.43	3	0.48	0.61	1	0.61	0.61	1	0.61
Male	125 / 80	5.5	Hyper and optimal	1.16	3	0.39	1.16	3	0.39	0.00	0	0.00	0.00	0	0.00
Male	134 / 82	5.6	Hyper and optimal	1.34	3	0.45	1.34	3	0.45	0.52	1	0.52	0.00	0	0.00
Female	115 / 77	5.5	Hyper and optimal	2.70	6	0.45	2.48	5	0.50	1.31	2	0.66	1.31	2	0.66
Female	127 / 73	5.7	Hyper and pre-diabetic	1.64	4	0.41	1.64	4	0.41	0.51	1	0.51	0.00	0	0.00
Female	112 / 71	5.2	Normal and optimal	2.57	7	0.37	2.06	4	0.52	1.24	2	0.62	1.24	2	0.62
Female	126 / 75	5.7	Hyper and pre-diabetic	1.13	3	0.38	1.13	3	0.38	0.00	0	0.00	0.00	0	0.00
Female	118 / 73		Normal and	1.13	3	0.38	1.13	3	0.38	0.00	0	0.00	0.00	0	0.00
Female	119 / 72	4.7	Normal and optimal	1.13	3	0.38	1.13	3	0.38	0.00	0	0.00	0.00	0	0.00
Female	143 / 93	5.4	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	120 / 69	5.1	Hyper and optimal	3.25	9	0.36	2.82	6	0.47	1.22	2	0.61	1.22	2	0.61

Female	112 / 76	5.9	Hyper and pre-diabetic	2.28	5	0.46	2.12	4	0.53	1.29	2	0.65	1.29	2	0.65
Male	123 / 78	5.7	Normal and pre-diabetic	1.47	4	0.37	1.47	4	0.37	0.00	0	0.00	0.00	0	0.00
Male	123 / 73	5.5	Normal and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	118 / 75	5.4	Normal and optimal	2.82	7	0.40	2.56	5	0.51	1.74	3	0.58	1.24	2	0.62
Male	151 / 95	6.1	Hyper and pre-diabetic	4.11	11	0.37	3.45	7	0.49	1.80	3	0.60	1.29	2	0.65
Male	123 / 74	5.9	Normal and pre-diabetic	2.09	5	0.42	1.93	4	0.48	1.11	2	0.55	0.00	0	0.00
Female	108 / 69	5.2	Normal and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	121 / 65	5.9	Normal and pre-diabetic	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	127 / 81	5.4	Hyper and optimal	1.96	7	0.28	1.26	3	0.42	0.00	0	0.00	0.00	0	0.00
Female	123 / 78	5	Hyper and optimal	2.89	7	0.41	2.55	5	0.51	1.24	2	0.62	1.24	2	0.62
Female	110 / 67	5.1	Normal and optimal	2.93	7	0.42	2.59	5	0.52	1.28	2	0.64	1.28	2	0.64
Female	112 / 74	5.2	Normal and optimal	1.78	5	0.36	1.44	3	0.48	0.62	1	0.62	0.62	1	0.62
Male	129 / 82	5.7	Hyper and pre-diabetic	2.73	7	0.39	2.47	5	0.49	1.24	2	0.62	1.24	2	0.62
Male	130 / 83	5	Hyper and optimal	3.21	8	0.40	2.94	6	0.49	1.24	2	0.62	1.24	2	0.62
Male	153 / 99	5.1	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	107 / 70	5	Normal and optimal	2.53	7	0.36	2.11	5	0.42	0.00	0	0.00	0.00	0	0.00
Female	146 / 85	5.4	Hyper and optimal	2.32	6	0.39	2.06	4	0.52	1.24	2	0.62	1.24	2	0.62
Male	144 / 85	7.4	Hyper and diabetic	2.57	7	0.37	2.06	4	0.52	1.24	2	0.62	1.24	2	0.62
Male	117 / 77	5.8	Normal and pre-diabetic	2.28	5	0.46	2.12	4	0.53	1.29	2	0.65	1.29	2	0.65
Male	125 / 76	5.5	Normal and optimal	2.14	6	0.36	1.43	3	0.48	0.61	1	0.61	0.61	1	0.61
Male	120 / 83	5.6	Hyper and optimal	4.11	11	0.37	3.33	7	0.48	1.71	3	0.57	1.20	2	0.60
Female	111 / 72	5.3	Normal and optimal	3.99	10	0.40	3.55	7	0.51	1.91	3	0.64	1.40	2	0.70
Female	122 / 63	5.4	Hyper and optimal	2.06	6	0.34	1.65	4	0.41	0.00	0	0.00	0.00	0	0.00
Female	121 / 69	5.3	Hyper and optimal	2.28	5	0.46	2.12	4	0.53	1.29	2	0.65	1.29	2	0.65
Female	139 / 82	5.7	Hyper and pre-diabetic	2.99	7	0.43	2.62	5	0.52	1.30	2	0.65	1.30	2	0.65
Female	116 / 70	5.6	Normal and optimal	1.56	5	0.31	1.26	3	0.42	0.00	0	0.00	0.00	0	0.00
Male	109 / 65	5.4	Normal and optimal	3.94	12	0.33	3.28	8	0.41	0.61	1	0.61	0.61	1	0.61
Male	127 / 81	5.9	Hyper and pre-diabetic	2.57	7	0.37	2.06	4	0.52	1.24	2	0.62	1.24	2	0.62
Male	125 / 77	5.4	Normal and optimal	1.28	3	0.43	1.28	3	0.43	0.00	0	0.00	0.00	0	0.00
Female	118 / 75	5.2	Hyper and optimal	2.66	7	0.38	2.25	5	0.45	0.60	1	0.60	0.60	1	0.60
Male	124 / 77	5.1	Normal and optimal	2.66	7	0.38	2.25	5	0.45	0.60	1	0.60	0.60	1	0.60
Male	121 / 74	5.2	Normal and optimal	2.28	5	0.46	2.12	4	0.53	1.29	2	0.65	1.29	2	0.65
Male	147 / 94	5.6	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	118 / 78	5.9	Normal and pre-diabetic	2.36	5	0.47	2.14	4	0.53	1.31	2	0.66	1.31	2	0.66
Male	125 / 82	5.7	Hyper and pre-diabetic	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	127 / 77	5.4	Normal and optimal	3.96	12	0.33	3.24	8	0.40	0.61	1	0.61	0.61	1	0.61
Male	135 / 88	5.4	Hyper and optimal	2.82	7	0.40	2.56	5	0.51	1.74	3	0.58	1.24	2	0.62
Male	121 / 85	6.6	Hyper and diabetic	1.30	5	0.26	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	124 / 73	5.3	Normal and optimal	2.24	5	0.45	1.99	4	0.50	1.17	2	0.59	0.66	1	0.66
Male	130 / 74	5.5	Hyper and optimal	2.66	8	0.33	2.18	5	0.44	0.53	1	0.53	0.00	0	0.00

Female	116 / 68	6.9	Normal and diabetic	3.62	9	0.40	3.19	6	0.53	1.96	3	0.65	1.96	3	0.65
Male	124 / 80	6.1	Hyper and pre-diabetic	2.53	7	0.36	2.11	5	0.42	0.00	0	0.00	0.00	0	0.00
Male	156 / 95	5.8	Hyper and pre-diabetic	1.81	6	0.30	1.22	3	0.41	0.00	0	0.00	0.00	0	0.00
Male	129 / 86	6	Hyper and pre-diabetic	4.54	14	0.32	3.38	8	0.42	1.06	2	0.53	0.00	0	0.00
Female	125 / 79	5.7	Hyper and pre-diabetic	2.89	7	0.41	2.55	5	0.51	1.24	2	0.62	1.24	2	0.62
Female	123 / 78	6	Hyper and pre-diabetic	3.67	9	0.41	3.24	6	0.54	2.01	3	0.67	2.01	3	0.67
Male	112 / 71	5.1	Normal and optimal	1.24	4	0.31	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Male	141 / 93	5.9	Hyper and pre-diabetic	3.99	10	0.40	3.55	7	0.51	1.91	3	0.64	1.40	2	0.70
Female	130 / 84	5.5	Hyper and optimal	5.20	13	0.40	4.72	10	0.47	2.38	4	0.60	1.29	2	0.64
Male	113 / 72	5.5	Normal and optimal	1.85	6	0.31	1.22	3	0.41	0.00	0	0.00	0.00	0	0.00
Male	128 / 77	6.3	Normal and pre-diabetic	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Male	125 / 77	5.6	Normal and optimal	2.82	7	0.40	2.56	5	0.51	1.74	3	0.58	1.24	2	0.62
Female	119 / 74	5.3	Normal and optimal	1.76	5	0.35	1.43	3	0.48	0.61	1	0.61	0.61	1	0.61
Female	135 / 74	5.3	Hyper and optimal	0.82	2	0.41	0.82	2	0.41	0.00	0	0.00	0.00	0	0.00
Female	104 / 66	5.3	Normal and optimal	1.76	5	0.35	1.43	3	0.48	0.61	1	0.61	0.61	1	0.61
Male	113 / 75	6.2	Normal and pre-diabetic	2.32	6	0.39	2.06	4	0.52	1.24	2	0.62	1.24	2	0.62
Female	114 / 67	5.1	Normal and optimal	3.98	13	0.31	2.83	7	0.40	0.51	1	0.51	0.00	0	0.00
Male	117 / 69	5.4	Normal and optimal	2.32	6	0.39	2.06	4	0.52	1.24	2	0.62	1.24	2	0.62

In this table, several identified and calculated phenotypical and genetic parameters for each participant are listed. The MutPred mutational load (Mut. load) and MutPred adjusted load (Adj. load) as well as the number of variants (Num of var) for each participant is given, when all variants are considered or when only variants with MutPred pathogenicity scores above a specified threshold (0.3, 0.5 or 0.6) are considered. 24 h ABPM, 24 hour ambulatory blood pressure monitoring. Normotensive and hypertensive classifications are based on 24 hour systolic and diastolic ambulatory blood pressure measurements with hypertensive participants being those with 24 h systolic ABPM above 130 mmHg for males, 125 mmHg for females, and/or 24 h diastolic ABPM above 80 mmHg for males, 75 mmHg for females. HbA1c, Glycated haemoglobin is measured to calculate the three-month average plasma glucose concentration. HbA1c measurements were used to classify participants as having either optimal blood glucose levels (HbA1c < 5.7%), being pre-diabetic (HbA1c from 5.7% to 6.4%) or diabetic (HbA1c > 6.4%). MutPred, a pathogenicity score assigned to a variant using the MutPred program (*mutpred.mutdb.org/about.html).*

Supplementary Table 2: Top twenty-five variants with the highest number of instances in the SABPA cohort data set

Variant	MutPred score	Number of instances in cohort total	Number of instances in haplogroup L participants	Number of instances in haplogroup MN participants	Number of instances on Phylotree	Present in macro-haplogroup L on Phylotree	Present in macro-haplogroup MN on Phylotree
4232C	0.609	87	78	8	6	Yes	Yes
5460A	0.505	58	52	6	43	Yes	Yes
4216C	0.611	39	0	39	13	Yes	Yes
4917G	0.628	25	1	24	6	No	**Yes**
11172G	0.630	21	21	0	3	Yes	Yes
14798C	0.609	14	0	14	5	No	**Yes**
14000A	0.552	11	9	2	4	Yes	Yes
14178C	0.516	11	11	0	5	Yes	Yes
13789C	0.576	10	10	0	2	Yes	Yes
13886C	0.571	8	8	0	1	**Yes**	No
15257A	0.785	8	0	8	3	No	**Yes**
9098C	0.685	8	0	8	5	Yes	Yes
13780G	0.606	7	0	7	1	No	**Yes**
15812A	0.518	7	1	6	3	Yes	Yes
8618C	0.641	7	7	0	4	Yes	Yes
10114C	0.603	6	6	0	1	**Yes**	No
10128A	0.609	6	6	0	1	**Yes**	No
15312C	0.530	6	6	0	1	**Yes**	No
15617A	0.714	6	3	3	3	Yes	Yes
6150A	0.676	6	6	0	2	Yes	Yes
6723A	0.653	6	6	0	1	**Yes**	No
7119A	0.520	6	6	0	2	Yes	Yes
13129T	0.850	5	5	0	2	Yes	Yes
14334T	0.501	5	0	5	0	No	No
3434G	0.577	5	5	0	3	Yes	Yes

In this table, a description is given for variants with MutPred scores above 0.5, that have the most number of instances in the SABPA cohort data set. The MutPred score is given for each variant. The number of instances in the entire cohort; in the haplogroup L participants; and in the haplogroup MN participants are also given. Finally, the number of times a variant appears as a haplogroup defining variant on Phylotree (*http://www.phylotree.org)* is given, and whether a variant defines haplogroups within macro-haplogroup L or macro-haplogroup MN. Variants that only define either macro-haplogroup L or macro-haplogroup MN lineages, are shown in bold. MutPred: a pathogenicity score assigned to a variant using the MutPred program (*mutpred.mutdb.org/about.html).*