Genetic Factors Modulating Mitochondrial DNA Copy Number

Rebecca Ruth Brennan

BSc (Hons), MRes

Doctor of Philosophy

Institute of Genetic Medicine

Faculty of Medical Sciences

Newcastle University

September 2017

Abstract

Mitochondria are dynamic organelles whose principal role is the generation of cellular energy (ATP) through oxidative phosphorylation (OXPHOS). 13 OXPHOS subunits are encoded by the mitochondrion's polyploid circular genome (mtDNA), and the nuclear genome (nDNA) encodes the remaining subunits as well as proteins required for mtDNA maintenance. In addition to mitochondrial number, mtDNA copy number (mtDNA CN) varies between cell and tissue type, depending on metabolic demand and baseline mtDNA quality, and ranges from hundreds to thousands of copies per cell. mtDNA CN is often linked to mitochondrial dysfunction and the ubiquity of mitochondria results in a broad spectrum of dysfunction and clinical phenotypes; ranging from primary mitochondrial disorders to complex diseases such as cancer, type 2 diabetes, and Parkinson's disease.

Given the variability in mtDNA between individuals, it is possible that mtDNA CN is influenced by secondary factors. I hypothesise that nDNA diversity is a major component of mtDNA variability between individuals and will test this hypothesis by conducting a genome wide association study (GWAS) in a large, European, asymptomatic cohort (>8000 individuals), comparing nDNA genotype to mtDNA copy-number as a QTL.

Peripheral blood mtDNA CN was correlated to array-based and imputed nDNA genotype in a two-stage QTL analysis, utilising three independent replicative cohorts: UKBS, Newcastle, and ALPAC. In addition the effect of potential confounding biological variables such as age, gender, blood count, and potential methodological confounders such as assay variation, technical and biological replicate numbers, and differences in genotype platform were all assessed and used to improve the GWAS analysis.

Individual cohort analysis identified nuclear gene *UNC13C* (Unc-13 Homolog C), two intergenic, and one intronic SNP, which is in close proximity to *PSMD3* (Proteasome 26S Subunit, Non-ATPase 3), to be genome wide significant (GWS) ($p< 1.00E-07$) in individual cohort analysis. However these hits could not be replicated in meta-analysis. mtDNA variant analysis in all three cohorts revealed that mtDNA SNPs *G5460A* and *G5046A*, which identify as mitochondrial haplogroup W, were significantly associated to a significant reduction in mtDNA CN. Furthermore, our work identified gender-specific genetic differences, which was supported by a

significant decrease in mtDNA CN in males with age, but not females, and significant changes in mtDNA CN relative to blood cell type and proportions highlighted the importance of regulating for cellular heterogeneity. Additionally, no difference in mtDNA CN was observed between pre- and post-menopausal women.

This work indicates that there are likely genetic variants present at the population level modulating mtDNA CN, but that this process is complex and multifaceted.

This thesis is dedicated to my family; to those who have passed and to those who are present. I hope this makes you proud.

Acknowledgements

First and foremost, I owe my sincerest thanks to my supervisors Dr Gavin Hudson, Dr Angela Pyle, and Professor Patrick Chinnery, for providing me with the opportunity to undertake this PhD project. I am extremely grateful for their continued support and guidance throughout my project. They have made my doctoral studies an enjoyable yet challenging experience, and I have learned a lot as a result of their supervision.

I would like to sincerely thank Dr Gavin Hudson for his patience in teaching me bioinformatics skills, and answering my never-ending ridiculous questions. I am continually grateful for all of your time, help, and perseverance to make the most out of my project. I would also like to thank Dr Angela Pyle for her ongoing caring, and supportive nature throughout my PhD. As well as managing my many laboratory technical problems, and teaching me how to do real-time PCR, Angela has also been my mentor. I am enormously grateful for your time and kind words which have kept me going during my hard times. I would also like to extend my thanks to Prof Patrick Chinnery for his continued efforts to enhance my PhD project, and for organising collaborations with University of Bristol.

I would also like to thank Prof Heather Cordell and Dr Joanna Elson for providing independent advice and guidance throughout my project, and I am grateful for the continued valuable feedback from everyone at the Wellcome Trust Centre for Mitochondrial Research, and the Institute of Genetic Medicine in Newcastle. I am particularly lucky to have had Hannah Lowes, Carla Roca, Siobhan Shuttleworth, Dr Michele Giunta, and Dr Hannah Steele as colleagues and friends during my studies. They have encouraged me during the difficult times, and have always provided me with plenty of daily entertainment. You made work more fun.

A special thanks goes to collaborators Anna Guyatt and Dr Santi Rodriguez at University of Bristol for their continued hard work, and enthusiasm. It has been a pleasure working with you both. Additionally, I am thankful for the funding provided by the Wellcome Trust and Barbour Foundation who sponsored my project.

I cannot thank my family enough for their unending encouragement, advice, humour, and love throughout my studies. My siblings Laura, Thomas, Ella, and Grace have been my backbone, and have supported me through every situation without hesitation, and have never failed to provide me with hilarious stories. My dad, Dr

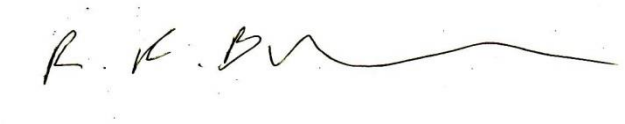
Vincent Brennan, has given me all of the opportunities to achieve all that I have, for which I am forever grateful. I am also very fortunate to have had my step-mum - Charlotte Brennan, and aunts – Dr Anne Astin and Dr Bryanie Shotton, provide invaluable advice and reassurance during my studies, and my little brothers Jack and Billy have continued to entertain me with their cheeky, and fun character.

I am eternally grateful for my amazing friends who have been an infinite source of encouragement, support, and joy during my PhD. In particular, Rebecca Claytor and Meggie Uren have always been there for me with brilliant advice and positive reinforcement, and Marie Bolster kindly provided me with delicious German chocolate during this write-up. Also, Dr Andy Smith has kept me sane with endless laughs from ridiculous antics, and has always provided reassurance, even from trips to the pub. I have many happy memories with him.

Finally, I owe my deepest thanks to my mum Annabel Brennan. Although not present, she taught me to believe in myself and to follow my dreams no matter how difficult it got.

Author's Declaration

This thesis is submitted for the degree of Doctor of Philosophy at Newcastle University. I, Rebecca Brennan, declare that the work described here is my own, unless where clearly acknowledged and stated otherwise. I certify that I have not submitted any of the material in this thesis for a degree qualification at this or any other university.



Rebecca R Brennan

Chapter-specific Declarations

All Results Chapters

United Kingdom Blood Service (UKBS) samples and individual demographic information were collected as part of the Wellcome Trust Case Control Consortium (WTCCC). Full blood count information for UKBS samples was provided by Professor Willem H Ouwehand at University of Cambridge as part of an ongoing collaboration with Professor Patrick Chinnery, now at University of Cambridge.

Results Chapters II and III

As part of the work in this thesis, a collaboration was organised between Professor Patrick Chinnery, now at University of Cambridge and Dr Santi Rodriguez at University of Bristol, whereby data about the UKBS cohort and the Avon Longitudinal Study of Parents and Children (ALSPAC) would be shared.

All mtDNA CN measurements in ALSPAC samples were performed by Dr Philip Guthrie, previously at University of Bristol. Nuclear DNA genotyping and imputation of ALSPAC samples was performed by Dr Gibran Hemani at University of Bristol, Dr George McMahon at University of Bristol, and Professor Jonathan Marchini at University of Oxford. Mitochondrial DNA genotyping of ALSPAC samples was performed by 23andme, the Wellcome Trust Sanger Institute, and the Laboratory Corporation of America. ALSPAC mitochondrial genotyping was cleaned, and haplogroup assignments conducted by Dr Mesut Erzurumluoglu as part of his PhD

with Dr Santi Rodriguez at University of Bristol. Estimated blood count information was generated for ALSPAC samples as part of the Accessible Resource for Integrated Epigenomics subset (ARIES) sub-study.

All genome-wide association analysis (GWAS), and further statistical analysis reported for ALSPAC samples in results chapters two and three were performed by Anna Guyatt, who was a member of Dr Santi Rodriguez's group at the University of Bristol. Anna Guyatt provided ALSPAC summary data to Newcastle University as part of the collaboration and work provided in this thesis.

Table of Contents

List of Figures

List of Tables

List of Publications

Pyle, A., Lowes, H., **Brennan, R**., Kurzawa-Akanbi, M., Yarnall, A., Burn, D. and Hudson, G. (2016) 'Reduced mitochondrial DNA is not a biomarker of depression in Parkinson's disease', Movement Disorders, 31(12), pp. 1923-1924.

Pyle, A., **Brennan, R**., Kurzawa-Akanbi, M., Yarnall, A., Thouin, A., Mollenhauer, B., Burn, D., Chinnery, P.F. and Hudson, G. (2015) 'Reduced cerebrospinal fluid mitochondrial DNA is a biomarker for early-stage Parkinson's disease', Annals of Neurology, 78(6), pp. 1000-1004.

Abbreviations

Δψm	Mitochondrial Membrane Potential
18S	18S Ribosomal Protein
ACL	ATP-citrate Lyase
AD	Alzheimer's disease
ADP	Adenosine Diphosphate
AG	Anna Guyatt (part of Dr Santi Rodriguez's group in the School of Social and Community Medicine at University of Bristol, UK)
AIDS	Acquired Immune Deficiency Syndrome
AKT	Protein Kinase B
ALS	Amyotrophic Lateral Sclerosis
ALSPAC	Avon Longitudinal Study of Parents and Children
AMP	Adenosine Monophosphate
AMPK	AMP-Activated Protein Kinase
ANT1	Adenine Nucleotide Translocator 1
ARIES	Accessible Resource for Integrated Epigenomics subset of ALSPAC
ATP	Adenosine Trisphosphate
B2M	Beta-2-microglobulin gene
BA	Absolute Basophil Count
β_G	Effect Size
BP	Basophil Proportion
bp	Base Pair/ Position
BN-PAGE	Blue Native Polyacrylamide Gel Electrophoresis

CAMKK2	Calcium-calmodulin-dependent Protein Kinase Kinase 2
ccf-mtDNA	Circulating, cell-free Mitochondrial DNA
COX	Cytochrome C Oxidase
CSF	Cerebrospinal Fluid
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleoside Triphosphates
ddPCR	Droplet-digital PCR
DGUOK	Deoxyguanosine Kinase
EA	Absolute Eosinophil Count
EDTA	Ethylenediaminetetraacetic acid
eIF	Eukaryotic Initiation Factor
ENDOG	Endonuclease G
EP	Eosinophil Proportion
ER	Endoplasmic Reticulum
ERs	Oestrogen Receptors
ERRα	Oestrogen-related Receptor alpha
FA	Fatty Acid
$FADH_2$	Flavin Adenine Dinucleotide
FBCs	Full Blood Counts
Fe-S	Iron-Sulphur
GAPDH	Glyceraldehyde 3-phosphate Dehydrogenase
GPx	Glutathione Peroxidase
GWAS	Genome-wide Association Study

GWS	Genome-wide Significant
H_2O_2	Hydrogen Peroxide
HB	Haemoglobin
HCT	Haematocrit
HPA	Hypothalamic Pituitary Adrenal gland
HRC	Haplotype Reference Consortium
HRT	Hormone Replacement Therapy
Hsp70	Heat Shock Protein 70
IBD	Identical by Descent
IGF	Insulin-like Growth Factor
IL	Interleukin
IMM	Inner Mitochondrial Membrane
INFO	Imputation Score
iPSC	Induced Pluripotent Stem Cell
ITT	Independent T-test
KO	Knock-out
LA	Absolute Lymphocyte Count
LHON	Leber’s Hereditary Optic Neuropathy
LincRNA	Long Intergenic Non-coding RNA
LP	Lymphocyte Proportion
LR-PCR	Long-range Polymerase Chain Reaction
MA	Absolute Monocyte Count
MAF	Minor Allele Frequency

MCH	Mean Corpuscular Haemoglobin
MCHC	Mean Corpuscular Haemoglobin Concentration
MCV	Mean Corpuscular Volume
MDS	Mitochondrial Depletion Syndromes
MDD	Major Depressive Disorder
MDV	Mitochondrial Derived Vesicles
MELAS	Mitochondrial Encephalopathy, Lactic Acidosis and Stroke-Like Episodes
MERRF	Myoclonus Epilepsy with Ragged Red Fibres
Mfn	Mitofusin
MIDD	Maternally Inherited Diabetes and Deafness
MNGIE	Mitochondrial Neurogastrointestinal Encephalopathy Disease
MP	Monocyte Proportion
mRNA	Messenger RNA
MS	Multiple Sclerosis
mtDNA	Mitochondrial DNA
mtDNA CN	Mitochondrial DNA Copy Number
mt-MRCA	Mitochondrial Most Recent Common Ancestor (Mitochondrial Eve)
MT-ND1	Mitochondrial NADH-ubiquinone Oxidoreductase Core subunit 1 gene
MT-ND4	Mitochondrial NADH-ubiquinone Oxidoreductase Core subunit 4 gene
MT-RNR2	Mitochondrial encoded 16S Ribosomal RNA
mtSNP	Mitochondrial SNPs

mtSSB	Mitochondrial Single-stranded DNA-binding Protein
mTOR	Mechanistic Target of Rapamycin
MWU	Mann Whitney-U test
NA	Absolute Neutrophil Count
NADH	Nicotinamide Adenine Dinucleotide
NCBI	National Centre for Biotechnology Information
nDNA	Nuclear DNA
NF-κB	Nuclear Factor kappa B
NP	Neutrophil Proportion
NRFs	Nuclear-Respiratory Factors
NUMT	Nuclear Mitochondrial DNA sequence
O_H	Origin of heavy (leading) strand replication
O_L	Origin of light (lagging) strand replication
OMM	Outer Mitochondrial Membrane
OXPHOS	Oxidative Phosphorylation
PAM	Presequence Translocase-Associated Motor
PBWT	Positional Burrows-Wheeler Transform
PCA	Principal Component Analysis/ Principle Components
PCR	Polymerase Chain Reaction
PD	Parkinson's disease
PEO1	*Twinkle* gene
PG	Dr Philip A.I. Guthrie (Dr Santi Rodriguez's group in the School of Social and Community Medicine at University of Bristol, UK).

PGC1α	Peroxisome Proliferator-Activated Receptor-gamma coactivator-1 alpha
PI3K	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PLT	Platelet
PMF	Proton Motive Force
POLG	Mitochondrial DNA Polymerase Gamma
PPAR	Peroxisome Proliferator-Activated Receptor
Q	Cochran's Q test
QC	Quality Control
qPCR	Quantitative Real-time Polymerase Chain Reaction
QQ	Quantile-quantile Plots
QTL	Quantitative Trait Locus
RA	Rheumatoid Arthritis
RBC	Red Blood Cell
rCRS	revised Cambridge Reference Sequence
RNA	Ribonucleic Acid
RNR	Ribonucleotide Reductase
ROS	Reactive Oxygen Species
RRF	Ragged Red Fibres
SAM	Sorting and Assembling Machinery
s.d	Standard Deviation
SDH	Succinate Dehydrogenase
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SE	Standard Error

SIRT	Sirtuins
SNP	Single Nucleotide Polymorphism
SOD	Superoxide Dismutase
StAR	STeroidogenic Acute Regulatory
T2DM	Type 2 Diabetes mellitus
TCA cycle	Tricarboxylic Acid Cycle
TFAM	Mitochondrial Transcription Factor A
TFB1M	Mitochondrial Transcription Factor B1
TFB2M	Mitochondrial Transcription Factor B2
TIM	Translocases of the Inner Mitochondrial Membrane
TK2	Mitochondrial Thymidine Kinase 2
TOM	Translocases of the Outer Mitochondrial Membrane
Twinkle	Mitochondrial Helicase
TYMP	Thymidine Phosphorylase
WB	Whole Blood
WBC	White Blood Cell
WGS	Whole Genome Sequencing
WTCCC	Wellcome Trust Case Control Consortium
Ubc	Ubiquitin C
UKBS	UK National Blood Service

Chapter 1 General Introduction

1.1 The Mitochondrial Organelle

Mitochondria are double membrane-bound organelles in eukaryotic cells which are fundamental to cellular maintenance. Described as the powerhouse of the cell, mitochondria generate over 80% of cellular energy in the form of adenosine trisphosphate (ATP). They also regulate calcium signalling, cell growth, and differentiation, cell cycle control, and cell death.

The mitochondrial structure consists of an outer membrane (OMM) and an inner membrane (IMM) separated by an intermembrane environment space (Figure 1.1). Mitochondria have their own matrix separate to that of the cellular cytoplasm, which provides an independent environment for mitochondrial-specific metabolic reactions to occur as well as mitochondrial DNA replication.

The OMM is porous and readily diffuses ions and uncharged molecules into the mitochondria through porins, or uses protein translocases to transport larger molecules (Weeber *et al.*, 2002). Evidence suggests that porins on the OMM are crucial for regulating mitochondrial calcium-induced calcium release, as the porins network with the adenine nucleotide transporter in the IMM and cyclophilin D in the mitochondrial matrix (Crompton *et al.*, 1998; Vyssokikh *et al.*, 2001).

The IMM on the other hand is non-porous, and all molecules and ions are transported into the mitochondrial matrix by specific membrane transport proteins (Kühlbrandt, 2015). The IMM forms cristae structures which increase surface area for aerobic respiration components to anchor (Schenkel and Bakovic, 2014). Both the OMM and IMM are made up of basic phospholipids, however phosphatidylglycerol and cardiolipin are specific to the mitochondria (Ren *et al.*, 2014).

The intermembrane space between the IMM and OMM is where the mitochondrial membrane potential ($\Delta\Psi$) is generated due to the constant flux of ions across the IMM. The $\Delta\Psi$ is usually around 180mV but is vital to mitochondrial maintenance and ATP generation (Kühlbrandt, 2015).

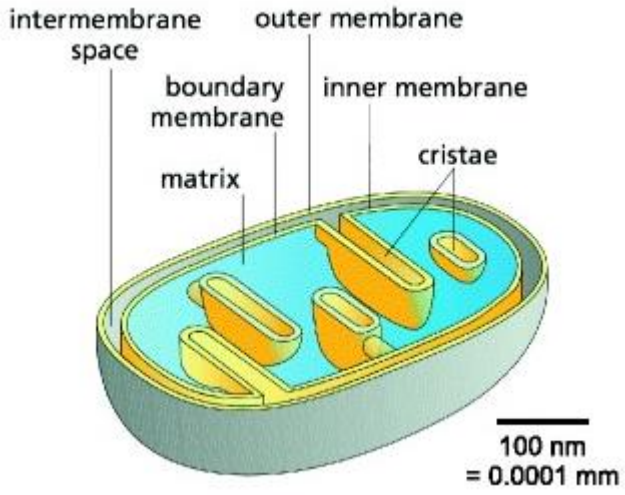


Figure 1.1: Membrane compartments in the mitochondrion. Image taken from (Kühlbrandt, 2015)

1.1.1 Mitochondria Evolution

Eukaryotic cells did not always contain mitochondria. Mitochondria are actually endosymbiotic remains and evolved from an invading α-proteobacterium billions of years ago (Nass and Nass, 1963a; Nass, 1966; Margulis, 1971; Lang *et al.*, 1997; Koonin, 2010), however uncertainty still remains. Despite the evidence of phylogenetic alignments suggesting this might be true (Ingman et al., 2000), Gray argues the majority of the mitochondrial proteome exists outside of the α-proteobacteria's evolutionary origin (Gray, 2012), and recent evidence suggests that the original host cell was already genetically chimeric before the mitochondrial symbiosis occurred. This suggests that mitochondria evolved later in eukaryotic evolution than was initially thought (Pittis and Gabaldón, 2016).

The smallest genomes recorded are in symbiotic bacteria in the stomachs of cicada (McCutcheon *et al.*, 2009), which evolved due to pressures exerted by intracellular conditions and competitions with other endosymbionts (Van Leuven *et al.*, 2014). Eukaryotic mitochondria also developed a smaller and more compact genome over time due to intracellular pressures, but many genes are believed to have translocated to the nuclear genome (Brown *et al.*, 1979; Lane and Martin, 2010b).

Beneficial genes and single-nucleotide polymorphisms (SNPs) have been conserved through evolution (Koonin, 2010), providing functional advantages to the cell under different pressures. The tricarboxylic-acid (TCA) cycle metabolic network has been the most highly conserved across species through evolution, demonstrating that the main function of the mitochondria is centred around the TCA cycle and oxidative phosphorylation (OXPHOS) activity (Chang *et al.*, 2010).

1.2 Metabolism and the Mitochondria

Mitochondria are present in multiple species, including prokaryotes, and bring a number of benefits to the cell. They store calcium, and synthesise haem, phospholipids, and cholesterol, and regulate cell apoptosis (Susin *et al.*, 1999; Duchen, 2000; Malik and Czajka, 2013). Mitochondria regulate cellular metabolic activity, playing a major role in cell homeostasis by generating energy in the form of ATP through OXPHOS. In eukaryotes, they are able to produced 36 molecules of ATP per glucose molecule metabolised through aerobic respiration, which is more beneficial than five ATP molecules generated through anaerobic respiration (Lin *et al.*, 2012; Müller *et al.*, 2012).

Additionally, mitochondria contribute to the anabolic and catabolic metabolism of carbohydrates, amino acids, and lipids - as they are home to the TCA, beta-oxidation, and many amino acid metabolism pathways (Figure 1.2).

1.2.1 Glycolysis and the TCA Cycle

Glycolysis is a metabolic pathway which occurs in the cytosol of eukaryotic cells. Through a series of enzyme reactions, glucose is broken down into pyruvate which is then transported into the mitochondria by the mitochondrial pyruvate carrier (Halestrap, 1975), and subsequently metabolised into either acetyl-coA by pyruvate dehydrogenase or oxaloacetate by pyruvate carboxylase. Both acetyl-coA and oxaloacetate feed in to the TCA cycle and regulate its flux, which is important for cellular metabolic equilibrium (Das *et al.*, 2015). During gluconeogenesis, lipogenesis, neurotransmitter biosynthesis, and glucose-induced insulin secretion, oxaloacetate is an important intermediate to drive TCA cycle (Jitrapakdee *et al.*, 2008). Acetyl-coA is vital for maintaining cellular respiration from glycolysis and is important for acetylating histones which control DNA expression (Wellen *et al.*, 2009; Lee *et al.*, 2014a).

1.2.2 Beta-oxidation

Pyruvate is a key substance which drives several intersecting pathways within the TCA cycle, including beta-oxidation (as reviewed by (Gray *et al.*, 2014)). Both beta-oxidation and the TCA cycle occur in the mitochondria (Mitchell, 1961), however beta-oxidation is the breakdown of dietary fatty acids (FAs) such as palmitic acid, oleic acid, and linoleic acid (Wanders *et al.*, 2010). During beta-oxidation, two-carbon

units are removed from the carboxyl end of the FA molecule to produce acetyl-CoA, which then feeds in to the TCA cycle, as reviewed by (Bartlett and Eaton, 2004).

1.2.3 Amino Acid Metabolism

Metabolism of amino acids is also imperative for metabolic homeostasis, however only six amino acids (leucine, isoleucine, valine, asparagine, aspartate, and glutamate) can be metabolised by mitochondria (Wagenmakers, 1998; Duan *et al.*, 2017). Leucine and part of isoleucine support glucose uptake, mitochondrial biogenesis, and beta-oxidation, whilst inhibiting protein degradation as they are oxidised to acetyl-CoA and utilised in the TCA cycle (Duan *et al.*, 2016). The carbon skeletons of the other amino acids are synthesised into glutamine or intermediates for the TCA cycle and also help maintain TCA cycle flux.

NADH and $FADH_2$ are reducing agents generated as bi-products during glycolysis, TCA cycle, beta-oxidation, and amino acid metabolism and are important metabolites as they power OXPHOS (Figure 1.3).

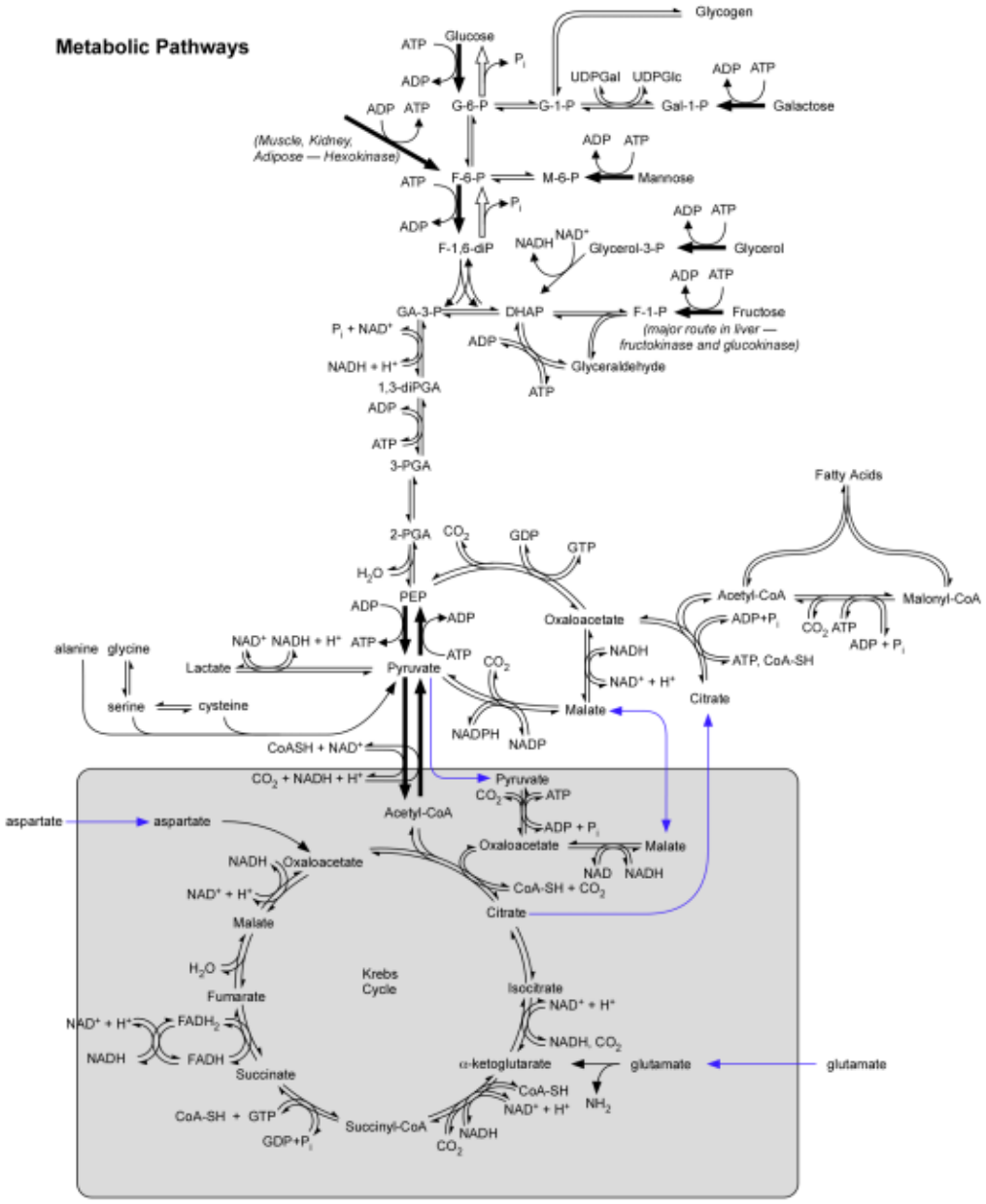


Figure 1.2: The biochemical reactions occurring during glycolysis, the tricarboxylic acid cycle, β oxidation and amino acid metabolism. Grey box indicates reactions occurring in mitochondria opposed to those in the cytoplasm. Image taken from http://www.indiana.edu/~oso/Fructose/MolBiol.html (Indiana University, 2014).

1.2.4 Oxidative Phosphorylation

On the surface of the mitochondrial IMM sit heteromultimeric structures of proteins responsible for OXPHOS. OXPHOS is an aerobic system comprised of five multi-subunit complexes, known as complexes I, II, III, IV, and V. The network is made up of over 70 subunits which are encoded for by the nuclear DNA (nDNA) (Pierron *et al.*, 2012). Complexes I to IV are responsible for the electron transport chain process, where electrons are used from NADH and are transferred across each complex through a series of chemical reactions, causing protons to be pumped out of the mitochondrial matrix into the intermembrane space, generating a proton motive force (PMF). Complex V is an ATP synthase responsible for chemiosmosis, which then uses the proton gradient generated to convert ADP in to ATP, which then is released into the cell (Figure 1.3).

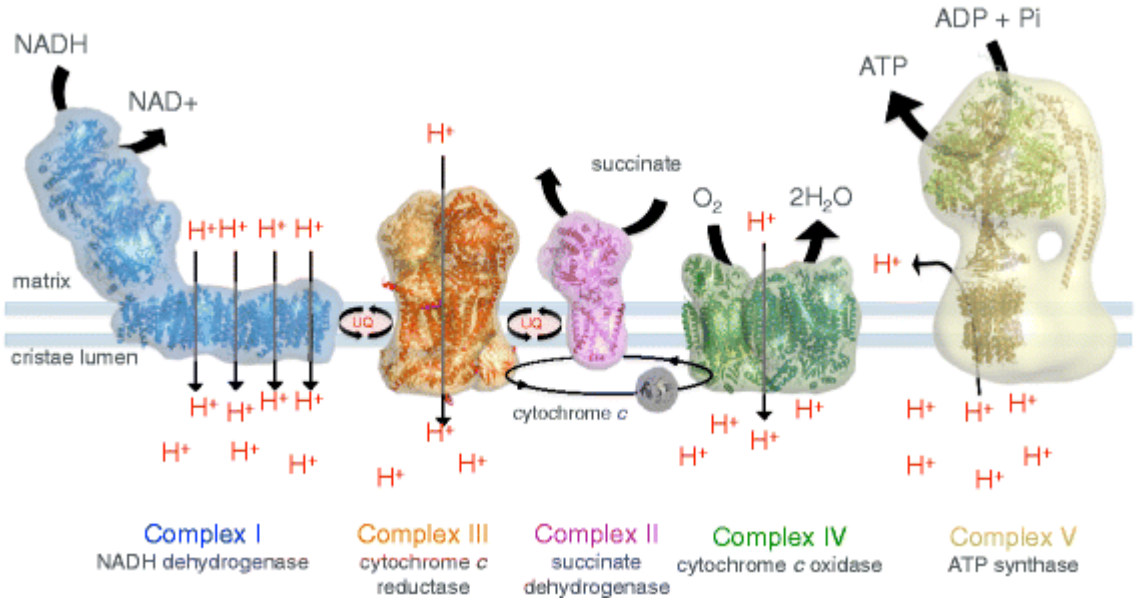


Figure 1.3: The five electron transport complexes of the oxidative phosphorylation respiratory chain. Blue: complex I (NADH/ubiquinone oxidoreductase), pink: complex II (succinate dehydrogenase), orange: complex III (cytochrome c reductase), green: complex IV (cytochrome c oxidase) and cream: complex V (ATP synthase). Through a series of reactions during oxidative phosphorylation electrons are transported across complexes. Complexes I, III and IV pump protons across the inner-mitochondrial membrane, generating a proton gradient which drives ATP synthesis. UQ means ubiquinol. Image taken from (Kühlbrandt, 2015).

1.2.4.1 Oxidative Phosphorylation and Disease

More than 150 distinct mitochondrial syndromes exist, and are predominantly caused by defects in the OXPHOS pathway (reviewed by (Leonard and Schapira, 2000; Wallace, 2005a). Changes to diet, and an imbalanced intake of carbohydrate, fat, and protein, introduce metabolic pressures which can alter mitochondrial metabolic sensitivity and lead to disease (Randle *et al.*, 1963; Sas *et al.*, 2016; Jørgensen *et al.*, 2017). However, defects in the human OXPHOS system are the biggest cause of

metabolic disorders, with a prevalence between one in 5000 and one in 8000 births (Skladal *et al.*, 2003; Thorburn, 2004; Jacobs *et al.*, 2006).

OXPHOS defects can be multisystemic or tissue-specific and predominantly affect high energy demanding tissues like muscle, brain, and the heart. Symptoms can also present at childhood or develop later in adulthood. Leigh Syndrome is considered the most typical example of an OXPHOS defect disease, where sufferers present with optic atrophy, hypotonia, ataxia, and dystonia before the age of two and often die a few years later. In addition to this disease, mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes (MELAS), and myoclonus epilepsy with ragged red fibres (MERRF) are considered 'classic' mitochondrial diseases. More detailed information about the basis of these, and other, mitochondrial diseases can be found here (Smeitink *et al.*, 2001; Thorburn, 2004; Jacobs *et al.*, 2006).

To diagnose mitochondrial disorders related to OXPHOS defects, skin or muscle biopsies are usually taken, and a series of biochemical and enzymology assays are performed to identify mitochondrial defects.

1.2.4.1.1 Biochemical Assays

Initial, minimally invasive tests can be performed in blood plasma, urine or cerebrospinal fluid (CSF) to identify the source of metabolic stress. Elevated lactate (cell redox state, normal <20) or pyruvate levels, as well as elevated amino acid levels such as alanine or proline, or TCA cycle intermediates, indicate inefficient aerobic respiration in mitochondrial disorders (Robinson, 2006; Haas *et al.*, 2008; Magner *et al.*, 2011). If initial tests suggest mitochondrial dysfunction, then skin or muscle biopsies are usually taken, and biochemical OXPHOS assays are performed. These tests assess the function of individual OXPHOS complexes, normally by directly measuring metabolite levels such as cytochrome c and coenzyme Q_1 (Thorburn *et al.*, 2004; Haas *et al.*, 2008).

1.2.4.1.2 Histology/ Histochemical Assays

Within each dysfunctional tissue, the quality and distribution of mitochondria are different. Histological and histochemical analyses of muscle or skin tissues are used to identify the severity of OXPHOS defects, and the presence of nuclear or mitochondrial-related genetic mutations. The traditional method for diagnosis involves using a modified Gomori trichrome stain to identify ragged red fibres (RRF), which stains proliferating mitochondria red in the subsarcolemmal space of muscle

fibre atrophy (Figure 1.4a). However, other histochemical assays are used to stain for NADH dehydrogenase, succinate dehydrogenase (SDH) and cytochrome c oxidase (COX) to assess complex I, II and IV activities of OXPHOS respectively (Figure 1.5b and c). The COX assay usually stains brown, and produces mosaic patterns for tRNA mutations and mitochondrial deletions (Figure 1.4c), whereas the SDH assay stains blue, and is useful to identify nDNA mutations, because all SDH subunits are coded for by nDNA (Figure 1.4b). The overlay of COX and SDH highlight the RRF staining (Bourgeois and Tarnopolsky, 2004; Tanji and Bonilla, 2007). Recently a quadruple immunofluorescence technique has been developed to detect complexes I and IV, whilst measuring porin to determine mitochondrial mass, as well as staining protein laminin to outline muscle cells (Grünewald *et al.*, 2014; Rocha *et al.*, 2015).

In addition to staining, sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and blue native polyacrylamide gel electrophoresis (BN-PAGE) are used in conjunction with Western Blotting to detect protein levels (Capaldi *et al.*, 2004; Schagger, 2006; Wittig *et al.*, 2006). Full details of procedures to diagnose mitochondrial disorders are detailed here (Jacobs *et al.*, 2006; Haas *et al.*, 2008).

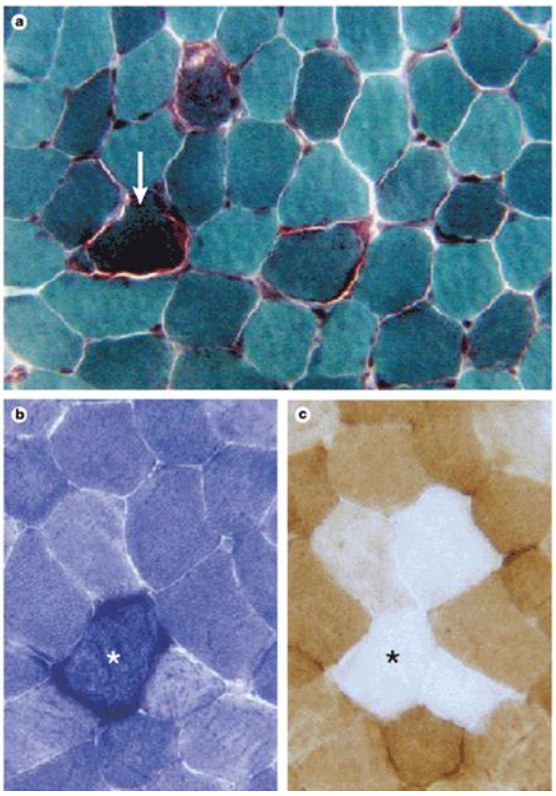


Figure 1.4: [A] Modified Gomori trichrome staining identifies ragged-red fibres – the red deposits around subsarcolemmal or intermyofibrillar space (arrow) which depicts abnormal mitochondria in muscle. In the RRF (asterisk), histochemical analysis using [B] succinate dehydrogenase and [C] cytochrome c staining identify enhanced [B] succinate dehydrogenase/ complex two activity and [C] no cytochrome c/ complex four activity. Image taken from (Smeitink et al., 2001).

Computer modelling has been used to predict the effects of enzyme deficiencies or defects in OXPHOS components on the TCA cycle and mitochondrial metabolic activity (Smith and Robinson, 2011; Zieliński *et al.*, 2016), and it is apparent that mitochondria are adaptable; however there is a threshold effect in tissues where accumulation of dysfunctional mitochondria result in a range of heterogeneous disease (Zeviani and Antozzi, 1997; Breuer *et al.*, 2013; Stewart and Chinnery, 2015)

OXPHOS defects are not responsible for all metabolic diseases though. Mitochondrial quality and biogenesis are also imperative, and both are controlled by genetic makeup.

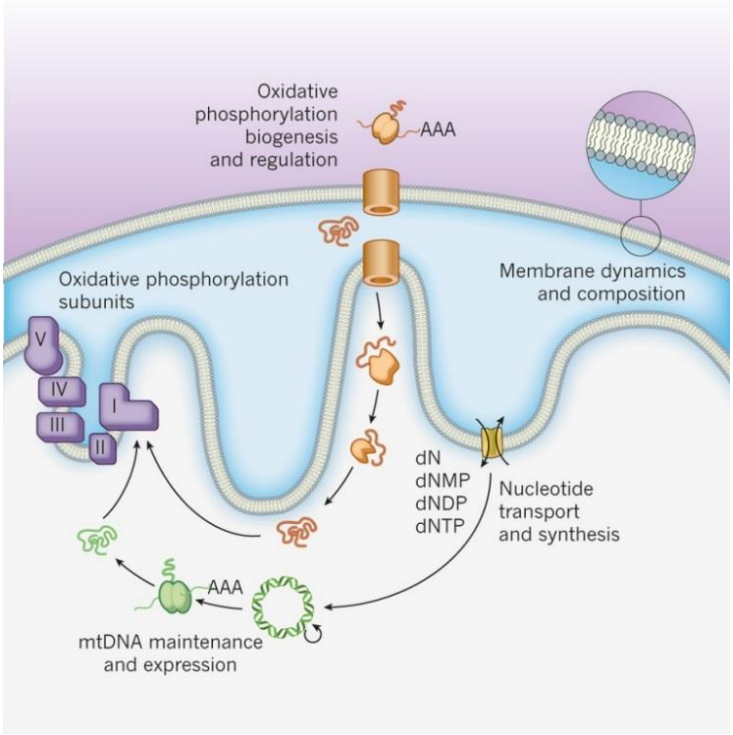


Figure 1.5: An overview of the interconnecting systems which regulate OXPHOS complex formation and function. Mitochondrial biogenesis and regulation, membrane dynamics and composition, nucleotide transport and mitochondrial DNA maintenance and expression all contribute. Image taken from (Vafai and Mootha, 2012).

1.3 Mitochondrial Quality Regulation and Biogenesis

1.3.1 Fusion and Fission in Mitochondria

Mitochondria are dynamic organelles and resemble their bacterial ancestors as they continually migrate bidirectionally around the nucleus using the cytoskeleton's microtubules and actin filament tracks (Hollenbeck and Saxton, 2005). Mitochondria continually undergo fusion and fission with one another (Chan, 2012), and interact with other cellular components like the endoplasmic reticulum (ER) (Friedman *et al.*,

2011; Lewis *et al.*, 2016), undergoing biogenesis and autophagy to regulate mitochondrial populations.

Mitochondrial morphology is dependent on cellular metabolic homeostasis, and mitochondrial mass is heterogeneous in each tissue, depending on their environmental demands (Rossignol *et al.*, 1999; Mishra *et al.*, 2014). For example, in skeletal-muscle mitochondria proficiently oxidise fatty acids, whereas in the brain they metabolise ketones, and in adrenal tissue mitochondria actively metabolise steroid hormones. Even within individual cells, mitochondria exhibit heterogeneity. In skeletal muscle, intermyofibrillar and subsarcolemmal mitochondria have different fuel sources and physiological reactions (Cogswell *et al.*, 1993).

Mitochondrial fusion usually occurs in cells with high metabolic demand, like muscle cells. Fusion allows bioenergetically efficient and electrically coupled systems to form, and respiratory machinery is shared to protect against high stress-induced energy demands or starvation (Figure 1.6) (Skulachev, 2001). Fusion uses GTPase machinery embedded in mitochondrial membranes and is controlled by mitofusins 1 (Mfn1) and 2 (Mfn2), and Opa1. Mfns are located on the OMM and Opa1 is located on the IMM. As well as promoting fusion, Opa1 regulates cristae structure which contributes to mitochondrial bioenergetic behaviour (Mishra *et al.*, 2014).

Fission segregates mitochondria during times of low metabolic activity, splitting networks to allow smaller mitochondrial units to be transported to other areas of cells where metabolic demand is higher (i.e. in neurons) (Westermann, 2010) (Figure 1.6). Mitochondrial division sites are marked by ER tubules circling and constricting mitochondria. These sites are believed to form as a result of changes in the membrane composition or because of calcium-influenced constriction regulation (Friedman *et al.*, 2011). Drp1 is an IMM dynamin-related protein which binds to the OMM, constricting and fragmenting mitochondria, causing them to undergo fission to become highly fragmented organelles (Westermann, 2012; Pyakurel *et al.*, 2014).

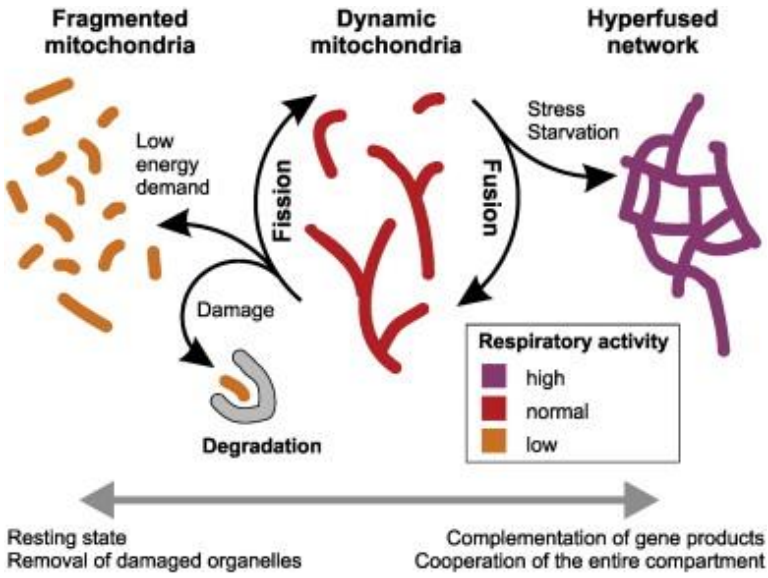


Figure 1.6: Schematic representation of mitochondrial morphological network changes when cellular metabolic demands change. At times where respiratory activity is low, mitochondria are fragmented. As respiratory activity increases, mitochondria fuse to form more dynamic and integrative networks. Image reproduced from (Westermann, 2012).

As a result of undergoing fission, some mitochondria may lack replication machinery or sufficient mitochondrial DNA (mtDNA), which can result in the loss of the mitochondrion's OXPHOS function, and loss of the membrane potential. This loss of membrane potential usually identifies damaged or poor quality mitochondria, and through the PINK1/Parkin signalling system, a mito-autophagosome forms around the mitochondria and they are selectively removed through a process known as mitophagy (Narendra *et al.*, 2010; Youle and Narendra, 2011).

Despite fission systems like PINK1/Parkin being important, especially in Parkinson's disease development (Narendra *et al.*, 2010; Pickrell and Youle, 2015), fusion machinery has been reported to be more important in managing mitochondrial biogenesis in yeast, *C. elegans* and differentiated muscle (Chen *et al.*, 2010; Hori *et al.*, 2011; Meyer and Bess, 2012). This is because when mitochondria, or mitochondrial DNA (mtDNA), are damaged, fusion events functionally compensate by sharing mtDNA and replication machinery, thereby maintaining replication and mitochondrial homeostasis. This suggests that the number of mtDNA are an additional, and important factor to manage mitochondrial biogenesis.

1.3.2 Mitochondrial Biogenesis

Mitochondrial maintenance is a fine balance between mitophagy and biogenesis, and is dependent on biochemical signalling and genetic response. Maintenance is especially important during differentiation (Westermann, 2010; Palikaras *et al.*, 2015)

and constant mitochondrial production is imperative because cell division relies on mitochondrial concentration (Jajoo *et al.*, 2016).

As previously described, mitochondria are biochemically adaptive organelles and respond to changing cellular energetics. In muscle, increased levels of ATP-citrate lyase (ACL) cause increased cardiolipin levels in mitochondria without changing cardiolipin synthesis gene expression levels, and improves OXPHOS complex formation and mitochondrial respiration (Das et al., 2015). ACL was activated by insulin-like growth factor-1 (IGF-1) through the PI3K/AKT pathway, which is known to control mitochondrial mass – pathway details are provided here (Egerman and Glass, 2014).

Sirtuins (SIRT), a family of nicotinamide adenine dinucleotide (NAD)-dependent deacetylase proteins, are also known to control mitochondrial biogenesis. SIRT1 is the most well studied sirtuin and is mostly present in the nucleus. SIRT1 has been known to deacetylate PGC-1α and regulate gene expression of many metabolic-related genes (Nemoto *et al.*, 2005; Rodgers and Puigserver, 2007). SIRT5 on the other hand is extensively located in the mitochondria, but is also regulated by PGC-1α (Buler *et al.*, 2014). SIRT5 has recently been found to directly bind cardiolipin and regulate OXPHOS activity (Zhang *et al.*, 2017). A more detailed overview of sirtuins' effects on mitochondrial biogenesis is reviewed here (Sack and Finkel, 2012).

Biochemical stimuli and genetic communication interconnect to oversee mitochondrial biogenesis. Mitochondria contain their own genome (mtDNA) which replicates in the mitochondrial matrix using about 80 nuclear-encoded replication proteins (Kenmochi *et al.*, 2001). Mitochondrial biogenesis however requires the co-ordinated import and regulation of around 1000-1500 nuclear encoded proteins, each being translated on cytoplasmic ribosomes before importation (O'Brien, 2002; Baker *et al.*, 2007). Some precursor proteins are synthesised to contain positively charged N-terminal presequences which transverse with the IMM and OMM. Sorting and assembling machinery (SAM) on the OMM inserts β-barrel proteins to allow precursors to translocase through the membrane with the aid of the mitochondrial membrane potential and heat-shock protein 70 (Hsp70). Hsp70 is regulated by presequence translocase-associated motor (PAM) to send precursors into the matrix. Precursor proteins are then cleaved by matrix proteases and folded in the mitochondrial matrix. Important translocases of the OMM or IMM (TOM) and (TIM)

respectively regulate where each protein is directed and are major components of biogenesis (Jornayvaz and Shulman, 2010).

The IMM is composed of boundary membranes, cristae, and cristae junctions. These subregions constantly remodel in response to stress and the cell's metabolic needs. Hoppins and colleagues, and von der Malsburg and colleagues, both discovered a large mitofilin/ Fcj1 protein complex on the IMM which acts as a scaffold for other proteins to assemble, regulating IMM morphology and organization. This complex also regulates importation of nuclear-encoded proteins needed to assemble IMS proteins, which is crucial for mitochondrial biogenesis (Hoppins *et al.*, 2011; von der Malsburg *et al.*, 2011).

Pickard and colleagues also show specialised inter-membrane junctions interact with dense mitochondrial cristae junctions for co-ordinated mitochondrial biogenesis (Picard *et al.*, 2015). This also involves a number of complex feedback systems between nDNA and mtDNA known as anterograde (nuclear to mitochondrial signalling), and retrograde (mitochondrial to nuclear) signalling.

McDermott-Roe and colleagues functionally demonstrated that endocnuclease G (ENDOG) in cardiomyocytes boosts mitochondrial biogenesis as well as stimulating mtDNA synthesis and transcription. ENDOG increases biogenesis because it is directly controlled by oestrogen-related receptor alpha (ERRα) and peroxisome proliferator-activated receptor (PPAR) -γ coactivator-1α (PGC1α) transcription factors, which are key components in mitochondrial biogenesis. Expression of ENDOG also clusters with other mitochondrial and metabolic genes, and McDermott-Roe *et al.* show ENDOG binding directly to mtDNA like mitochondrial transcription factor A (TFAM) (McDermott-Roe *et al.*, 2011). ENDOG has also been seen to directly cleave nDNA and stimulate apoptosis (Li *et al.*, 2001).

ERRα is one member of a subfamily of nuclear receptors which are known to control expression of metabolic genes involved in glucose and glutamine metabolism, mitochondrial activity, lipid handling, and energy sensing. ERRs prefer to bind close to promoters of genes and work independently of oestrogen signalling, however are still important for biogenesis (reviewed by (Audet-walsh and Giguere, 2015)). In comparison, oestrogen receptors (ERs) have more strict binding domains and require the presence of oestrogen to regulate biogenesis (Chen *et al.*, 2009).

Unlike ERs, ERRs are dependent on the presence of PGC1-α and PGC1-β (Gaillard *et al.*, 2006; Gaillard *et al.*, 2007). PGC-1α binds PPARs (which are also known to control lipid metabolism and sensing (Berger and Moller, 2002)), and activates a number of transcription factors including nuclear-respiratory factors (NRFs) and TFAM, which regulate mitochondrial replication and transcription (Virbasius and Scarpulla, 1994). PGC-1α was first discovered in adipose tissue determining the switch between white or brown adipose tissue through adaptive thermogenesis. This process involves increased mitochondrial biogenesis and uncoupling of respiration (Puigserver and Spiegelman, 2003). PGC-1α is activated by the catalytic AMP-activated protein kinase (AMPK) alpha subunit via NRF-1 binding. AMPK is activated when intracellular levels of ATP decline and intracellular levels of AMP increase, such as during nutrient deprivation or hypoxia (Jornayvaz and Shulman, 2010). Mitochondrial biogenesis is not regulated by one sole pathway and Figure 1.7 gives an overview of the complex network of signalling molecules working together to balance both biogenesis as well as mitochondrial quality (Qi and Ding, 2012).

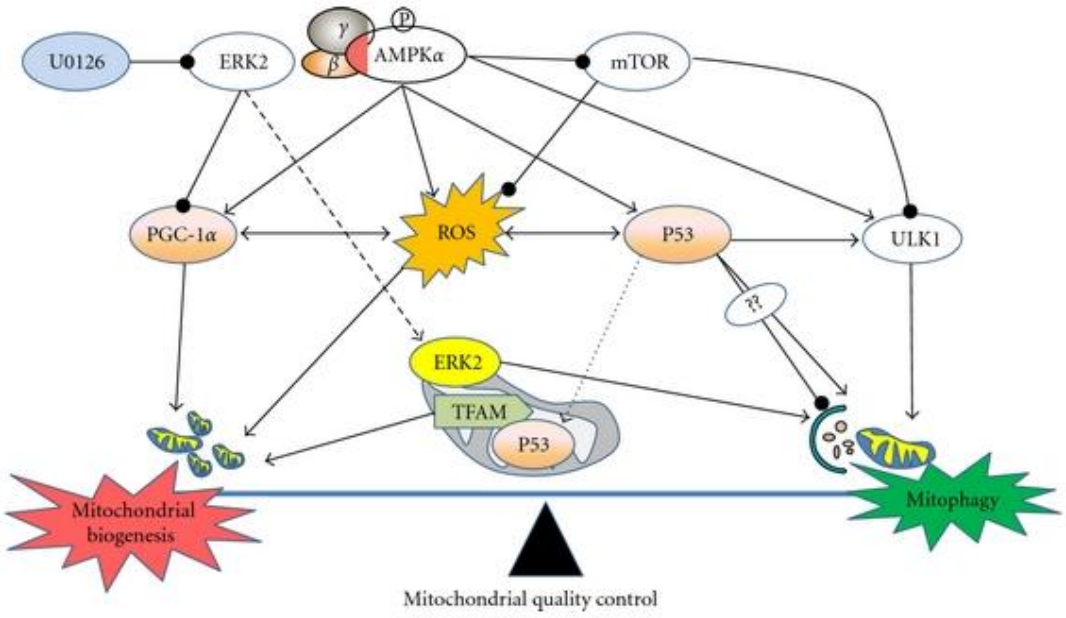


Figure 1.7: Mitochondrial quality control is a fine balance between biogenesis and mitophagy, where both regulate overall mitochondrial numbers in a positive or negative manner respectively. Biogenesis and mitophagy occur to meet cellular metabolic demands. Activated AMPK activates mammalian Atg-1 homologues (ULK1) to prevent mitophagy and trigger mitochondrial biogenesis via PGC-1α-dependent transcription. When nutrients are available, mTOR represses mitochondrial biogenesis and ULK1-dependent mitophagy, thus affecting the net balance of defective mitochondria with new functional mitochondria. AMPK: AMP-activated protein kinase; mTOR: mammalian target of rapamycin; PGC-1α: PPARgamma coactivator 1-alpha; ULK1: the mammalian Atg1 homologs, uncoordinated family member (unc)-51: like kinase 1; ERK2: the extracellular signal-regulated protein kinase 2; U0126: ERK inhibitor. Reproduced from (Qi and Ding, 2012).

Like ERs and ERRs, mechanistic target of rapamycin (mTOR) also promotes mitochondrial function by selectively translating nuclear-encoded mitochondrial messenger RNAs (mRNA) by inhibiting the translation of initiation factor 4E (eIF4E)-binding proteins (Ramanathan and Schreiber, 2009). As a serine/threonine kinase, mTOR has been evolutionarily conserved and responds to a number of stimuli, reviewed by (Morita *et al.*, 2013), inhibits autophagy, and is important for anabolic processes like cellular growth. mTOR is active only when glucose is available (Figure 1.8) (Chaube and Bhat, 2016). When glucose is limited, AMPK is activated and other nutrients like fat and amino acids are metabolised to generate ATP instead. This compensatory mechanism is beneficial to maintaining both mitochondrial and cellular homeostasis, but it is also detrimental when trying to treat cancer, as it contributes to tumorigenesis. This summarises the complex interplay between nutrient and genetic regulation of mitochondrial biogenesis pathways.

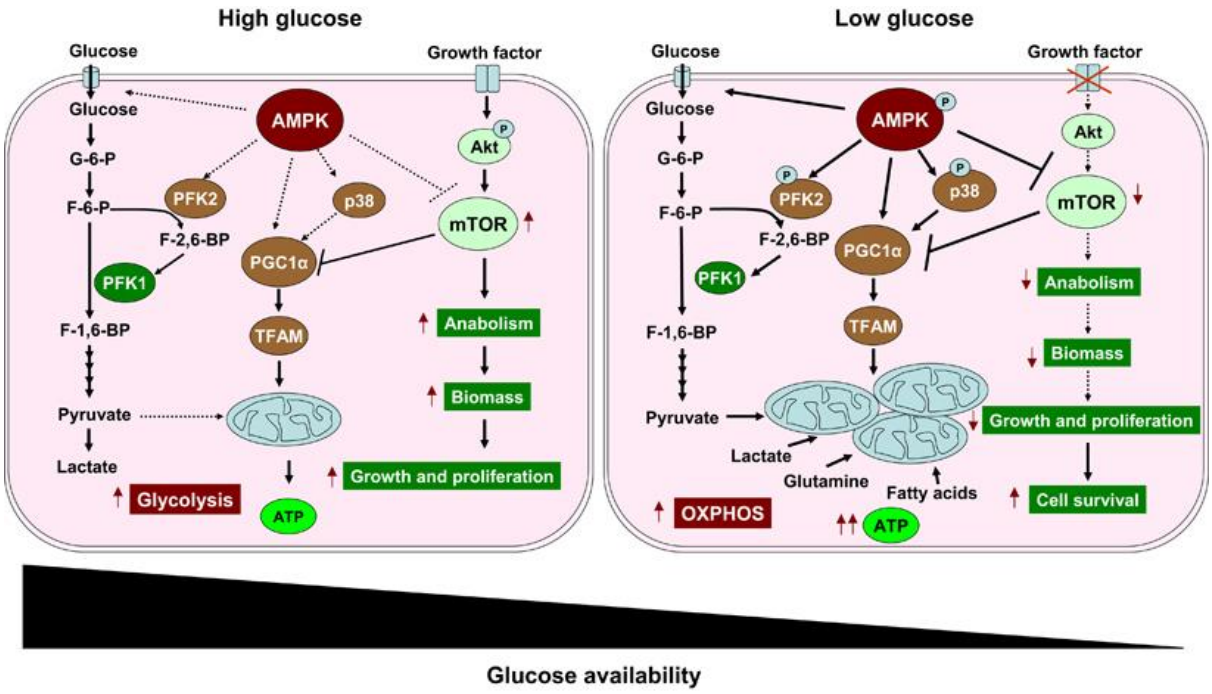


Figure 1.8: AMPK regulation of mitochondrial biogenesis under different glucose conditions. When glucose is readily available, AMPK is inactive and mTOR is activated, causing mitochondrial biogenesis and cell proliferation. In glucose deprived conditions, AMPK is activated and mTOR suppressed causing other nutrients like fats and amino acids to be oxidatively metabolised to generate ATP. Image taken from (Chaube and Bhat, 2016).

Radzvilavicius and colleagues used mathematical modelling to argue that the complexity behind regulating mitochondrial biogenesis and overall quality is what lead to eukaryotes having different gamete producing cells compared to prokaryotes, as prokaryotes are less likely to encounter problems and accumulate mutant or defective mitochondria (Radzvilavicius *et al.*, 2016).

1.4 Mitochondrial DNA

Reminiscent to their ancestors, mitochondria are unique organelles as they still contain their own genome (mtDNA). mtDNA is a double-stranded, circular molecule made up of a heavy and a light strand. The heavy strand has a higher molecular weight than the light strand due to differences in composition of purines (adenine and guanine) and pyrimidines (thymine and cytosine). The heavy strand contains more guanine than the light strand (Stewart and Chinnery, 2015).

mtDNA was first sequenced in 1981 by a group led by Fred Sanger at University of Cambridge (Anderson *et al.*, 1981) and was later revised in 1999 by a group at Newcastle University (Andrews *et al.*, 1999). The revised Cambridge Reference Sequence (rCRS) of human mtDNA is now available online (www.mitomap.org).

mtDNA consists of 16,569 bp which code for 37 mitochondrial genes, consisting of 13 OXPHOS subunit-coding polypeptides, translated from 2 dicistronic and 11 monocistronic mRNAs (Jones *et al.*, 2008). The genome also codes for two ribosomal-RNAs (12S and 16S), 22 transfer-RNAs, and a non-coding D loop region (Anderson *et al.*, 1981; Taylor and Turnbull, 2005; Malik and Czajka, 2013) (Figure 1.9).

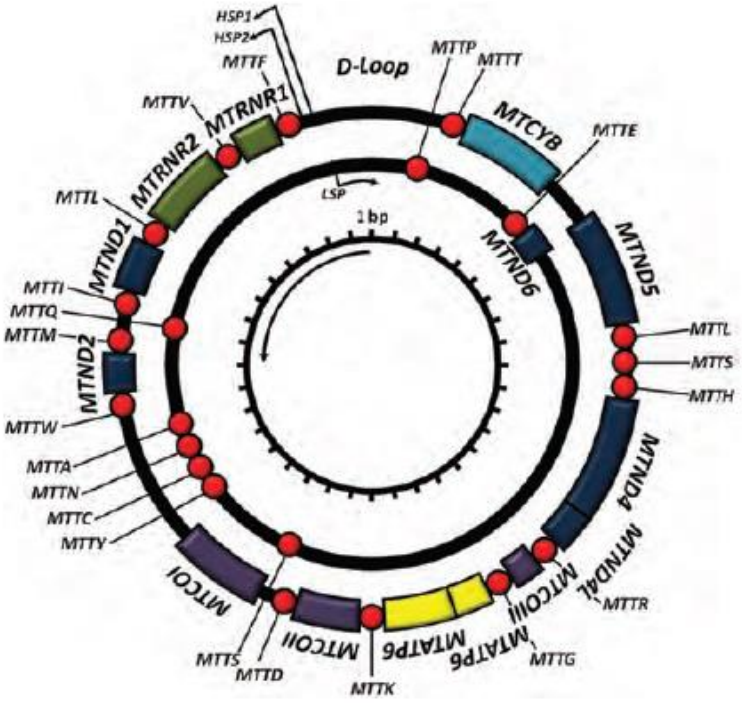


Figure 1.9: Schematic image of the mitochondrial genome and each respective region of coding or control (D-loop) DNA. Image taken from (Chinnery and Hudson, 2013).

mtDNA is highly conserved (Brown *et al.*, 1979; Yang *et al.*, 2014) and is inherited uniparentally down the maternal line (Hutchison *et al.*, 1974; Giles *et al.*, 1980), as paternal mtDNA is engulfed by maternal autophagosomes then degraded by lysosomes soon after oocyte fertilisation (Al Rawi *et al.*, 2011; Sato and Sato, 2011).

During human migration, mtDNA developed stable polymorphisms which now define mtDNA as specific haplogroups (van Oven and Kayser, 2009). These haplogroups form complex tree-and-branch like structures and are detailed in an online phylotree (http://www.phylotree.org/).

1.4.1 Haplogroups

Haplogroups are defined as evolutionary inherited mtDNA mutations which have created stable population subgroups, but are separated by common sequence variation (Chinnery and Hudson, 2013). Haplogroups were thought to have arisen over 10,000 years ago from our ancestors who have had to adapt to new diets and environments after migrating from our origin in Africa (Mishmar *et al.*, 2003; Cerezo *et al.*, 2012; Pierron *et al.*, 2012). In Africa, haplogroup L is most prevalent, whilst in Asia haplogroup M is the most common. In European populations, haplogroups H, V, J, T, M, I, X, W, U, and K are prevalent for over 95% of individuals (Torroni *et al.*, 1996).

Each haplogroup has its own unique set of stable polymorphisms present in the coding region of the mitochondrial genome (Torroni *et al.*, 1996), and contribute to both assembly of the respiratory chain (Pello *et al.*, 2008), and the bioenergetic behaviour of mitochondria (Suissa *et al.*, 2009; Hulgan *et al.*, 2011). For example, haplogroup Uk has a strong effect on OXPHOS and changes energy output (Gomez-Duran *et al.*, 2010). Haplogroup variability and bioefficiency has also been shown to effect ageing and regulate mitophagy (Harman, 1956; De Benedictis *et al.*, 2000; Gitschlag *et al.*, 2016), and mitochondrial tolerance contributes to cell survival over time (Ballard *et al.*, 2007).

Knowing this, specific haplogroups have been identified to be susceptible to developing certain metabolic diseases. Haplogroup J has been seen to modulate the penetrance of Leber's hereditary optic neuropathy (LHON) (Torroni *et al.*, 1997; Rose *et al.*, 2001; Hudson *et al.*, 2007b), and haplogroups H and V are more susceptible to Parkinson's disease (PD) (Pyle *et al.*, 2005; Hudson *et al.*, 2013a; Soto-Hermida *et al.*, 2014), whereas haplogroup Uk increases Alzheimer's disease (AD) risk (Lakatos

et al., 2010). On the other hand, haplogroups J, T, and K are seen to be protective against PD and osteoarthritis (Van Der Walt *et al.*, 2003; Ghezzi *et al.*, 2005; Hudson *et al.*, 2013b), and haplogroups U and T protect against AD (Shoffner *et al.*, 1993; Chagnon *et al.*, 1999; Van Der Walt *et al.*, 2004). What's more, haplogroup J influences the progression of Aquired Immune Deficiency Syndrome (AIDS) whereas haplogroups UK and N* reduce that progression (Hendrickson *et al.*, 2008). Hulgan's team also demonstrated that mitochondrial genomics influence metabolic physiology in AIDS patients taking certain medications (Hulgan *et al.*, 2011).

1.4.2 Mitochondrial DNA Replication

Like mitochondrial biogenesis, mtDNA replication is tightly regulated through a number of pathways to ensure mitochondria can accommodate to the changing metabolic demands in individual cells and tissues.

mtDNA is packaged into spherical shaped nucleoids with a diameter ranging between 70 to 100 nm, along with a mixed array of replication proteins - TFAM, mitochondrial single-stranded DNA-binding protein (mtSSB), polymerase-gamma (POLG), SUV3L, SHMT, and DHX30 (Bogenhagen *et al.*, 2008; Ylikallio *et al.*, 2010). TFAM and mtSSB both act as replication and packaging enzymes (Takamatsu *et al.*, 2002). Nucleoids constantly change shape and distribution when mitochondria undergo fusion and fission but they are continually tethered to the IMM where replication occurs.

The D-loop is the only non-coding region within mtDNA and contains the two transcription initiator sites P_H and P_L, as well as the heavy-strand replication origin (O_H) (Kelly and Scarpulla, 2004). The D-loop forms a triple-stranded structure approximately 700 bp from the O_H and is thought to regulate replication (Takamatsu *et al.*, 2002). Both the heavy and light strand of mtDNA have individual replication origin sites (Montoya *et al.*, 1982), however priming of mitochondrial replication occurs at the O_H promotor on the light chain (Chang and Clayton, 1985), causing the heavy chain to be replicated first.

Replication is believed to be conducted in an asymmetric manner, although a simultaneous model has been suggested (Holt *et al.*, 2000). In the asymmetric model (Clayton, 1982), transcription is initiated at P_L producing mRNA. At the three evolutionarily conserved blocks (CSB) I, II, and III, a decision is made to both cleave RNA and start DNA replication at O_H, or continue with transcription (Clayton, 1987;

Kelly and Scarpulla, 2004) (Figure 1.10). TFAM, mtSSB, and POLG bind to mtDNA to initiate replication, and TFAM binds four mtDNA sites (two within the promotor origin and two downstream) (Suissa *et al.*, 2009) which bends mtDNA into a U shape structure to allow the catalytic alpha-subunit of POLG to access the nucleotides and replicate them (Ropp and Copeland, 1996; Hallberg and Larsson, 2011). There is controversy about how much TFAM there is per mtDNA molecule - some suggest TFAM is up to 1700 fold higher than mtDNA (Takamatsu *et al.*, 2002), whereas others suggest there is insufficient TFAM per molecule (Bogenhagen *et al.*, 2008). Most genes being transcribed have polyadenylated tails added to signal stop codons. It is suggested that ribosomes recognise single-stranded start codons or structural sequences on the mRNA to translate them (Jones *et al.*, 2008).

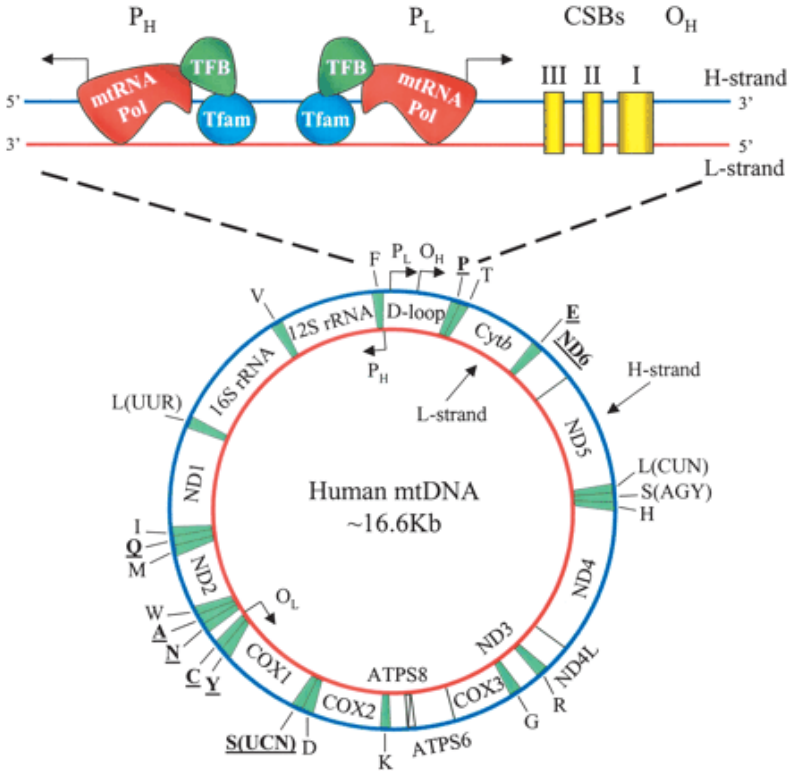


Figure 1.10: An overview of the 37 genes and replication sites within mitochondrial DNA. The expanded section shows only the non-coding D loop containing the transcription initiation sites PL and PH and the heavy chain replication origin OH. Replication machinery polymerase gamma (POLG), transcription factor A (TFAM) and B (TFB) bind at PL to start transcription. RNA is synthesised until the machinery passes the evolutionarily conserved blocks (CSBs) where a decision is made to either continue transcription or change to DNA synthesis at OH. The blue line represents the heavy strand and the red line represents the light strand where replication is asymmetric in a bidirectional manner. Reproduced from (Kelly and Scarpulla, 2004).

1.4.2.1 Consequences of Errors in Mitochondrial DNA Replication

mtDNA replication is independent to cell cycle nDNA replication and mtDNA avoids recombination (Elson *et al.*, 2001). mtDNA is more susceptible to mutations than nDNA because mitochondria lack protective histones, lack DNA repair mechanisms,

and compartmentalise DNA in the presence of reactive oxygen species (ROS), causing oxidative damage (Wallace *et al.*, 1987).

ROS are predominantly generated at complexes I and III during OXPHOS and are usually neutralised by superoxide dismutase (SOD), catalase, glutathione (GPx), thioredoxin and protein thiol systems. However, errors in OXPHOS are known to cause an indirect imbalance of nucleotide pools, TCA-cycle flux, one-carbon cycle, and ROS signalling, as the PMF is directly altered (Vafai and Mootha, 2012). As a result, the downstream effect of affecting nucleotide recycling directly affects mitochondrial replication, and errors in nucleotide pooling can lead to the development of mitochondrial deletions, or mitochondrial depletion (Nikkanen *et al.*, 2016). The lack of DNA protection or repair machinery also contributes to mtDNA errors. It is predicted that 1 in every 200 women carries a mitochondrial mutation (Reddy *et al.*, 2015).

As a result of exposure to so many possible errors during transcription, defects in the mtDNA result in a variety of mitochondrial diseases, with characteristics varying from neurological disorders, to muscular weakness, and endocrine dysfunction. Mitochondrial diseases arise from large-scale deletions or single point mutations in mtDNA. Large-scale deletions are usually sporadic but severe and occur during repair or replication of the genome (Krishnan *et al.*, 2008), causing diseases such as Pearson's syndrome or Kearns Sayre Syndrome (Russell and Turnbull, 2014). Single point mutations on the other hand generate a greater degree and variety of clinical presentations in individuals depending on the mutation and prevalence of it in a given tissue. Mitochondrial myopathy and infantile onset spinocerebellar ataxia are mitochondrial depletion syndromes which develop as a result of mitochondrial deoxyribonucleoside triphosphate (dNTP) pooling defects (El-Hattab and Scaglia, 2013; Nikkanen *et al.*, 2016).

Mitochondrial deletions and depletion syndromes are known to develop because of defects in nDNA autosomal genes, such as *POLG*, *PEO1,* and *TFAM*, which are involved in the maintenance of mtDNA. Thus emphasizing the importance of correct mtDNA: nDNA communication (Larsson *et al.*, 1998; He *et al.*, 2002; Nikkanen *et al.*, 2016).

1.5 Nuclear-Mitochondrial Intergenic Communication

1.5.1 Evolution of Nuclear-Mitochondrial DNA

During the evolutionary stages of the symbiote with the host cell, the majority of the original mtDNA migrated and incorporated into the nDNA in a process known as numtogenesis. Numtogenesis is thought to occur when double-stranded breaks develop in nDNA, and mitochondria close to the nucleus integrate their mtDNA to stabilise nDNA and prevent deletions (Hazkani-Covo and Covo, 2008). As a result, numtogenesis has been seen in more than 85 eukaryotic genomes, forming copies of the mtDNA in the host nDNA - known as nuclear-mitochondrial DNA segments (NUMTs) (Hazkani-Covo *et al.*, 2010). In humans, approximately 64-100% of germline NUMTs have mtDNA identities, and have been evolutionarily conserved for tens of millions of years (Mourier *et al.*, 2001; Dayama *et al.*, 2014). NUMTs are not fully understood in disease, but it has been reported that there are more NUMTs in cancer cells than normal cells, and are thought to alter mitochondrial biogenesis (Srinivasainagendra *et al.*, 2017).

Rickettsia prowazekii is an intracellular parasite and is thought to be more closely related to mitochondria than any other microbe because both genomes code for TCA cycle and OXPHOS components, and report nuclear gene homologues in their nDNA (Andersson *et al.*, 1998; Abhishek *et al.*, 2011). Allen suggests that some genes are retained in the functional organelles' DNA, like mitochondria, because they provide a reference location for nuclear-encoded genes to integrate to, facilitating rapid and direct regulatory coupling (Allen, 2003).

Stoycheva and colleagues also showed that yeast retrotransposons, such as *Ty1,* integrated from mtDNA into nDNA depending on mitochondrial function. With mitochondrial haplogroups developing during human migration, it has been suggested that endosymbiotic relationships between nDNA and mtDNA are different, and are dependent on an individual's mitochondrial biogenesis (Stoycheva *et al.*, 2007; Pierron *et al.*, 2012).

Using gel retardation, Ozawa's group also identified that during transcription of specific OXPHOS genes, the same nuclear protein factors bound to similar upstream sequences in both mtDNA and nDNA, and thus suggested mitochondrial biogenesis is finely regulated at a transcriptional level (Suzuki *et al.*, 1991). Taanman's group however argued otherwise, as mtDNA concentration is specific for a given tissue,

depending on its metabolic demands and therefore is not comparable (Van den Bogert *et al.*, 1993). These studies, combined with genomic studies which identify that NUMTs are different in size and continually edited, emphasise the complexity and uncertainty in nDNA and mtDNA intergenic communication (Albayrak *et al.*, 2016; Hazkani-Covo and Martin, 2017).

1.5.2 The Importance of Nuclear-Mitochondrial DNA Communication for Mitochondrial Maintenance

Mitochondrial maintenance requires finely coordinated communication between mtDNA and nDNA and relies on over 70 nuclear-encoded genes. The complex intricacies between the two genomes give rise to metabolic variation between individuals (Graham, 2012).

Several key nDNA encoded genes are known to be important for mtDNA regulation. *TFAM*, mitochondrial transcription factors B1 and B2 (*TFB1M* and *TFB2M*), *POLG*, mitochondrial helicase (*Twinkle* or *C10orf2*), *mtSSB*, and adenine nucleotide translocator 1 (*ANT1*) are a few well-characterised genes involved in mtDNA maintenance (Kaukonen *et al.*, 2000; Spelbrink *et al.*, 2001; Van Goethem *et al.*, 2001; Kang *et al.*, 2007; Ruhanen *et al.*, 2010; Harvey *et al.*, 2011; Müller-Höcker *et al.*, 2011; Ngo *et al.*, 2011; Milenkovic *et al.*, 2013; Rajala *et al.*, 2014).

1.5.2.1 Mitochondrial Transcription Factors A, B1, and B2 (TFAM, TFB1M, and TFB2M)

TFAM is a mtDNA binding protein that regulates packaging, expression and distribution of mtDNA, regulating both mtDNA replication and transcription (Lezza, 2012). Kaufman's group show TFAM binds multiple mtDNA as homodimers and compacts them in to nucleoids by bending the DNA backbone (Kaufman *et al.*, 2007; Ngo *et al.*, 2011). They suggest that TFAM alone is sufficient to organise mtDNA, however recent studies oppose this and show TFB2M is vital for initiation of transcription as it induces promotor melting, and TFAM alone is insufficient (Posse and Gustafsson, 2017; Ramachandran *et al.*, 2017). TFAM is upregulated simultaneously with mtDNA amount in COX deficient cells, acting as a compensatory mechanism for OXPHOS demands (Konokhova *et al.*, 2016). On the contrary, knocking out *TFAM* in mice causes mitochondria to become stressed and depletes mtDNA (Larsson *et al.*, 1998; Ekstrand *et al.*, 2004).

1.5.2.2 Polymerase Gamma (POLG)

POLG is also crucial for mtDNA maintenance. POLG is the enzyme responsible for mtDNA replication and is comprised of a catalytic subunit (encoded by *POLG*), and an accessory subunit (encoded by *POLG2*) (Olson *et al.*, 1995; Ropp and Copeland, 1996). Increases in the catalytic subunit have not shown to increase mtDNA levels, however increasing the accessory subunit does (Lefai *et al.*, 2000; Spelbrink *et al.*, 2000), suggesting the accessory subunit is more important (Carling *et al.*, 2011). Defects in *POLG* result in a range of mitochondrial diseases (Van Goethem *et al.*, 2001; Jacobs *et al.*, 2006).

1.5.2.3 Mitochondrial Helicase (Twinkle) and Single Stranded Binding Protein (mtSSB)

Both Twinkle and mtSSB improve the productivity of POLG (Farr *et al.*, 1999; Ruhanen *et al.*, 2010). Twinkle is the hexameric mtDNA helicase coded for by *PEO1* (Spelbrink *et al.*, 2001), and mtSSB is a tetrameric single-stranded binding protein coded for by *SSBP1*. Twinkle binds to nucleoid membranes and recruits mtDNA to the IMM where it unwinds it for replication (Milenkovic *et al.*, 2013; Rajala *et al.*, 2014), whereas mtSSB protects single-stranded DNA and promotes replication protein-protein interactions (Ruhanen *et al.*, 2010). Loss of Twinkle in mice, or mtSSB in flies and yeast results in mtDNA depletion, as Twinkle is predicted to hydrolyse dNTPs which are needed to replenish nucleotide pools (Ruhanen *et al.*, 2010; Milenkovic *et al.*, 2013; Nikkanen *et al.*, 2016).

1.5.2.4 Nucleotide Cycling Factors

To prevent depletion, efficient mtDNA replication also relies on having sufficient pools of nucleotides. Two pathways regulate these pools – the cytosolic import pathway, which imports nucleotides made by cytosolic ribonucleotide reductase (RNR) into the mitochondria, and the mitochondrial salvage pathway, which converts nucleosides into nucleotides inside the mitochondria. Mutations in cytosolic import pathway genes *TYMP* (thymidine phosphorylase) and *RRMB2B* (alternative subunit to RNR), and mitochondrial salvage pathway genes *TK2* (thymidine kinase 2) and *DGUOK* (deoxyguanosine kinase), all result in mitochondrial depletion syndromes (MDS) with varying phenotypes (Smeitink *et al.*, 2001; Limongelli and Tiranti, 2002; Zeviani and Di Donato, 2004; Carling *et al.*, 2011).

1.5.2.5 Other Genetic Regulators

NRF1 and NRF2 have also been shown to actively regulate mtDNA, as NRF1 directly activates TFAM and binds PGC1 which also controls mtDNA maintenance (Virbasius and Scarpulla, 1994; Wu *et al.*, 1999) as it induces approximately 600 genes whose products are mitochondrial related (Pagliarini *et al.*, 2008).

ANT1 codes for the protein which is a core structural element of the mitochondrial permeability transition pore and controls nucleotide concentrations in the cytoplasm by controlling nucleo-cytoplasmic signalling, which regulates energy consumption to OXPHOS. Mendelian mutations in *ANT1* cause a rare, autosomal dominant mitochondrial-related disorder called progressive external ophthalmoplegia (Kaukonen *et al.*, 2000; Limongelli and Tiranti, 2002).

Changes in mitochondrial morphology also affect cellular homeostasis and a highly co-ordinated regulation system is required to maintain mitochondrial number and overall mitochondria quality. Suppression of mic60, a morphological mitochondrial protein which interacts with mtDNA and regulates D-loop structuring, has been seen to reduce mitochondrial transcription and OXPHOS by preventing TFAM and mitochondrial RNA polymerase from binding mtDNA (Yang *et al.*, 2015). Furthermore, even though TFAM is required for mtDNA segregation (Kasashima *et al.*, 2011), the ER has been seen to orchestrate where mtDNA replication occurs and how it is distributed to daughter cells (Lewis *et al.*, 2016).

Additionally, mitochondrial haplogroups have been seen to modulate mitochondrial transcription and replication. Haplogroup J for example has increased transcription compared to haplogroup H, because haplogroup J variants increase TFAM binding (Suissa *et al.*, 2009).

All these factors emphasise the co-ordinated interactions between mtDNA and nDNA to manage mitochondrial maintenance. Genetic variability gives rise to differences in mitochondrial quality as well as mtDNA levels.

1.6 Mitochondrial DNA Copy Number

Mitochondrial DNA copy number (mtDNA CN) is central to mitochondrial vitality, affecting respiratory chain function (Robin and Wong, 1988; Wallace, 1999; Capps *et al.*, 2003; Gu *et al.*, 2013; Fukuoh *et al.*, 2014) and mitochondrial biogenesis (Suzuki *et al.*, 1991; Westermann, 2012; St. John, 2016). mtDNA codes for several structural OXPHOS subunits, therefore mitochondrial function and survival rely on adequate molecules of mtDNA to sustain OXPHOS capacity and maintain cellular homeostasis (Shoubridge and Wai, 2007; Van Blerkom, 2011). Changes in mtDNA CN can contribute to mitochondrial dysfunction, causing disease (Wallace, 1999; Silva *et al.*, 2000; Reinecke *et al.*, 2009; Wallace, 2012; El-Hattab and Scaglia, 2013; Giordano *et al.*, 2014; Pyle *et al.*, 2015a).

Each nucleoid contains 1-10 copies of mtDNA (Garrido *et al.*, 2003; Bogenhagen *et al.*, 2008) and cells can contain tens to hundreds of nucleoids, resulting in hundreds to thousands of copies of mtDNA being present in a single cell. The number of copies of mtDNA is known as mtDNA CN. mtDNA CN within each mitochondria is dependent on the genomic composition within each nucleoid; where an increased proportion of heteroplasmic mtDNA results in an increased number of nucleoids per mitochondrion (reviewed by Gilkerson, 2009).

1.6.1 Mitochondrial DNA Copy Number Regulation

mtDNA CN is highly balanced between synthesis and degradation for cellular maintenance. Despite key regulatory mechanisms not being known, a threshold hypothesis (Figure 1.11) has been suggested as a means for mtDNA CN regulation (Clay Montier *et al.*, 2009), however mediator mechanisms are still unknown.

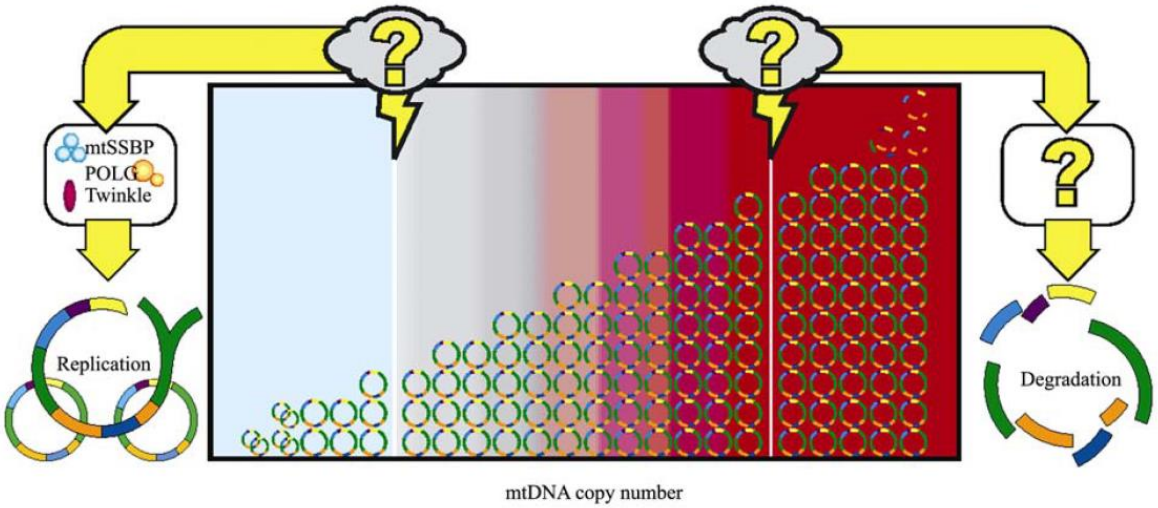


Figure 1.11: Mitochondrial DNA copy number (mtDNA CN) threshold hypothesis, where mtDNA CN is thought to be regulated by thresholds. The lower the threshold of unknown factors triggers up regulation and activation of replication machinery mitochondrial single-stranded binding protein (mtSSBP), polymerase-gamma (POLG) and replicative helicase (Twinkle). Whilst higher thresholds result in the activation of unknown factors which lead to mitochondrial DNA degradation. Reproduced from (Clay Montier et al., 2009).

Much of the literature has investigated biological factors directly regulating mtDNA CN, and amongst the complex signalling networks, protein quality, biochemical signalling, nutrient availability, and genetic expression all contribute to overall mtDNA CN regulation. However, no study to date has comprehensively investigated genetic variants causing mtDNA CN variation.

1.6.1.1 Protein Quality

Many of the functioning proteins required to maintain efficient mitochondrial and nuclear functions are iron-sulphur containing (Fe-S) clusters. Fe-S clusters predominantly form in the mitochondria, and are vital for mitochondrial function by regulating iron homeostasis. Once assembled, Fe-S clusters are then translated into Fe-S proteins in the cytoplasm by binding to cysteinyl sulphurs of the associated proteins. Enzymes involved in the formation of Fe-S clusters are highly conserved across species, from bacteria to humans (Beinert *et al.*, 1997). Fe-S clusters contribute to overall genome stability, both in the mitochondria and the nucleus, as they make up the majority of the replication, and maintenance proteins like DNA polymerases, primases, helicases, and glycosylases (reviewed by (Rouault and Tong, 2005; Paul and Lill, 2015)). Knowing this, it poses a confounding factor if protein quality might contribute to overall mtDNA CN regulation.

1.6.1.2 Biochemical Signalling and Stress Interference

Calcium is involved in differentiation, especially during development. Variation in calcium concentration play a key role in mitochondrial biogenesis changes by activating the AMPK pathway through calcium-calmodulin-dependent protein kinase kinase 2 (CAMKK2) (Hawley *et al.*, 2005; Woods *et al.*, 2005). AMPK is activated when intracellular levels of ATP decline and intracellular levels of AMP increase, such as during nutrient deprivation or hypoxia. This is more commonly seen in ageing and increased ROS generation, resulting in decreased mtDNA CN levels (Gomes *et al.*, 2013). This may be due to a secondary outcome as ROS homeostasis relies on the PMF to regenerate mitochondrial NADPH, and PMF drives ATP synthesis. PMF also indirectly controls biosynthetic reactions such as Fe-S cluster biogenesis and protein import (Vafai and Mootha, 2012). With age, OXPHOS efficiency decreases and contributes to more ROS production, therefore acting as a vicious cycle to mtDNA CN depletion with age (Fannin *et al.*, 1999; Payne and Chinnery, 2015).

1.6.1.3 Nutrient Availability

mtDNA CN fluctuates in a number of metabolic diseases such as cancer, type 2 diabetes mellitus (T2DM), obesity, and neurological disorders like PD. mtDNA CN was seen to directly negatively correlate to visceral fat accumulation in young adults (Lee *et al.*, 2014b), and improved diet has been able to correct for this (Hernández-Ríos *et al.*, 2013). Lipid availability has been shown to affect mitochondrial biogenesis, clustering, and bioenergetics (Haemmerle *et al.*, 2011; Kim and Park, 2015), and has been known to affect the quality of mitochondrial membrane formation, thus affecting overall quality of mitochondria. This affects their metabolic capacity and impacts on signalling, resulting in mtDNA CN changes.

Pyruvate metabolism has also been shown to contribute to OXPHOS efficiency and overall ATP production throughout the cell, and diseases such as cancer, PD, AD, and Leigh syndrome are all linked to specific enzyme deficiencies (Gray *et al.*, 2014). In PD alone, PGC1-α genes, which are known to regulate pyruvate metabolism are under expressed, and overexpression of *PGC1-α* in mice has been shown to prevent dopaminergic neuron loss (Zheng *et al.*, 2010). Furthermore, treatment with pyruvate in PD has been shown to alleviate ROS in dopaminergic cells, as pyruvate is a scavenger of ROS (Brand, 1997; Varma and Hegde, 2007). In addition to this, amino acids also contribute greatly to mitochondrial maintenance, where some of their

metabolites support immunological signalling events (reviewed by (Wu, 2009) and are known to contribute to disease development. Research is currently turning its attention to the human microbiome and its relationship with the nervous system (Sampson *et al.*, 2016; Hill-Burns *et al.*, 2017), suggesting nutrition may have an important influence on mtDNA CN regulation respectively.

Furthermore, resveratrol has been shown to increase mitochondrial biogenesis by activating SIRT1 (Viscomi *et al.*, 2011), and is suggested to be a good alternative therapy to stimulate biogenesis in those with defective mitochondrial function (Poole *et al.*, 2015).

1.6.1.4 Genetics

Transcription of nuclear-encoded OXPHOS components are mediated by Nrf1 or Nrf2 and PGC1-α is the key mediator of mitochondrial biogenesis, increasing respiration in muscle cells by interacting with uncoupling protein-2 (UCP2) as well as interacting with transcription factors such as PPARs, thyroid hormone, glucocorticoid, oestrogen, and oestrogen-related receptors (ERRs) (Jornayvaz and Shulman, 2010; Nargund *et al.*, 2015). PPAR pathway has also been involved in T2DM, inflammation, demyelination, and cancer (Berger and Moller, 2002). Thyroid hormone receptors which are activated by PGC1- α increase the transcription of mitochondrial genes (Weitzel *et al.*, 2003).

As well as PGC-1α, disrupting *SIRT-1* expression rescues septic mice from death (Liu *et al.*, 2015b). Pyle A. and colleagues showed mtDNA CN decreased in neutrophils in blood of septic patients (Pyle *et al.*, 2010), suggesting SIRT-1 may also be a critical regulator of mtDNA CN.

Stem cells especially require fine tuning, as their metabolic demands change during development and differentiation, resulting in continual dynamic shifts in mitochondrial biogenesis to accommodate to cellular stress (Harvey *et al.*, 2011). Liver kinase B *(Lkb1)* plays an essential role in haematopoietic stem cell maintenance and progenitor outcome. *Lkb1* regulates haematopoiesis via an mTORC1-independent pathway and E2F, Nrf1, and PPAR-γ motifs as the most significantly enriched promoter binding elements in the *Lkb1* pathway. *Lkb1* regulates PGC-1 expression and *Lkb1* knockout (KO) resulted in decreased PGC-1 expression and decreased mitochondrial function. Surprisingly, they also saw an increase in mtDNA CN over time, as a compensatory mechanism (Gan *et al.*, 2010).

Mitochondrial replication machinery like POLG, mtSSB, and TFAM are well characterised intermediates known to regulate mtDNA CN, however this is not the complete picture. *TFAM* overexpression directly increases mtDNA CN, and KO causes depletion. Despite these studies, the ratio of TFAM molecules to mtDNA CN is variable, ranging from one molecule per 10 to 1000 bp (Ekstrand *et al.*, 2004; Carling *et al.*, 2011). *TFAM* mRNA expression has also been seen to actively regulate the expression of sarco/ endoplasmic reticulum calcium ATPase 2 (*SERCA2*), which is involved in the uptake of calcium in cardiomyocytes, by binding directly to its promoter region on nDNA. mtDNA CN was directly proportional to *TFAM* mRNA levels (Watanabe *et al.*, 2011), however Suarez and colleagues showed calcium and TFAM alone are not responsible for mtDNA CN, as increased mtDNA CN did not decrease despite diminished *SERCA2* expression when cells were over exposed to glucose (Suarez *et al.*, 2008). mtSSB has been shown to correlate to mtDNA CN (Schultz *et al.*, 1998).

POLG is located on chromosome 15q24 and codes for a 139.5kDa protein (Ropp and Copeland, 1996). Mutations in *POLG* results in more mtDNA deletions in neurons, with increased mtDNA CN, and enhanced OXPHOS. This suggests POLG only regulates the quality of the DNA sequence but not overall mtDNA CN (Perier *et al.*, 2013b). During differentiation, methylation of *POLG* at exon 2 decreased mtDNA CN expression, introducing epigenetic controls on mtDNA CN (Kelly *et al.*, 2012; Lee *et al.*, 2015). Kelly and colleagues support this idea by demonstrating, using different mouse tissues, that mtDNA CN could be manipulated by altering DNA methylation of *POLG-α* (Kelly *et al.*, 2012).

Further to this, mitochondrial-nuclear interactions have been seen to fluctuate during the cell cycle in yeast (Grand *et al.*, 2014). Padovan-Merhar and colleagues used single-molecule multicolour mRNA fluorescence in situ hybridization to quantify the abundance of mRNA to cell volume in single-cell fibroblasts, and demonstrate that transcription activity can directly and globally be modulated by cellular volume, irrespective of nDNA content (Padovan-Merhar *et al.*, 2014). This suggests mtDNA CN regulation may be regulated by physical pressures within the cell that have not been well studied. Figure 1.12 gives an overview of the known complex feedback systems overseeing mitochondrial maintenance and mtDNA CN changes.

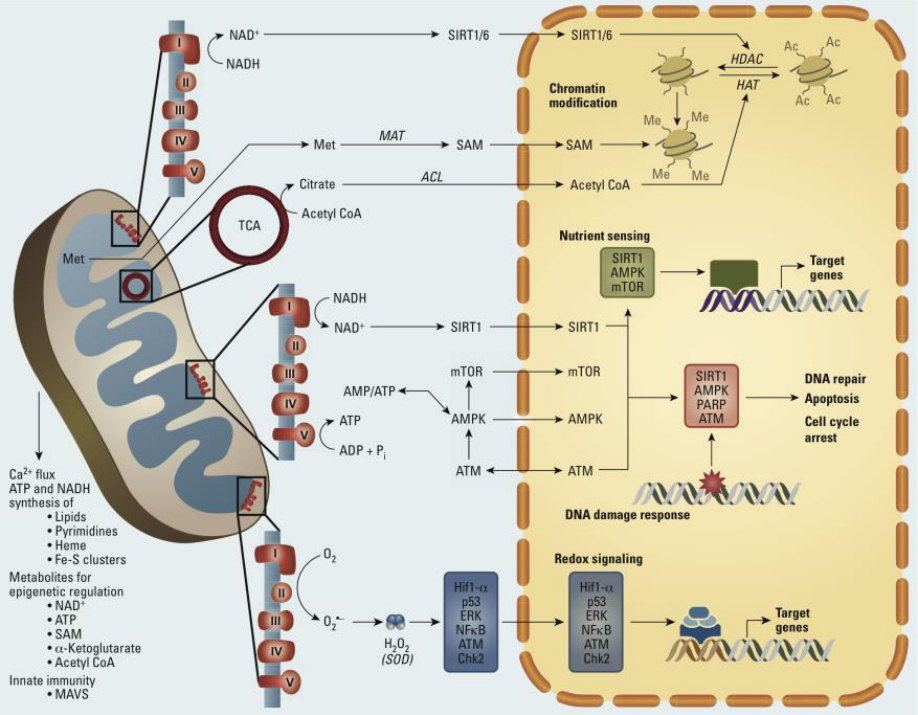


Figure 1.12: An overview of the complex systems feeding mitochondrial and nuclear genomic communication regulating overall cellular integrity. Byproducts like ATP and NADH from the TCA cycle and OXPHOS contribute to epigenetic modifications of nuclear DNA, as well as being involved in gene expression and DNA repair mechanisms via the sirtuin, mTOR or AMPK pathways. Oxygen radicals contribute to redox signalling and gene expression controlled by p53 and NF-kB pathways. In addition to this, mitochondria synthesise macromolecules and flux calcium as well as generating MAVS. Abbreviations: ACL, ATP citrate lyase; AMPK, AMP-activated protein kinase; Acetyl CoA - acetyl coenzyme A; ERK - extracellular signal-regulated kinase; HATs - histone acetyltransferases; HDACs - histone deacetylases; Hif1-α - hypoxia inducible factor 1 alpha; MAT - methionine adenosyltransferase; MAVS - mitochondrial antiviral signalling protein; Met - methionine; mTOR - mechanistic target of rapamycin; NFκB - nuclear factor kappa B; PARP - poly(ADP-ribose) polymerase 1; SAM - S-adenosylmethionine; TCA - tricarboxylic acid cycle. Reproduced from (Shaughnessy et al., 2014).

1.6.2 Homoplasmy and Heteroplasmy

Primary mitochondrial diseases, i.e. those caused by a defect in mtDNA, do not follow a Mendelian pattern of inheritance, as they are inherited down the maternal line (Giles *et al.*, 1980; Al Rawi *et al.*, 2011). The severity and range of tissues affected by mitochondrial disease is very broad and can be individual to each sufferer. This is often dependent on the mitochondrial mtDNA heteroplasmy levels in each tissue. Tissues can contain cells which are both homoplasmic and heteroplasmic. Homoplasmy is when all of the mtDNA within a cell is identical and heteroplasmy is when cells contain wildtype and mutant mtDNA (Holt *et al.*, 1990). Each tissue has a phenotypic threshold before biochemical manifestations result in clinical presentations (Chinnery *et al.*, 1997; Chinnery *et al.*, 2001; Gilkerson, 2009; Stewart and Chinnery, 2015). Thresholds are usually reported to be in the range of 60-95% (Russell and Turnbull, 2014).

As mitochondria contribute so significantly to the metabolic requirements within cells, highly metabolically demanding cells and tissues are more prone to mitochondrial disease and the development of neurological, cardiological, and muscular disorders (Taylor and Turnbull, 2005). For example, neurons are more likely to accumulate mutations over time because they rarely replicate and have a high metabolic demand which leads to higher production of ROS through oxidative stress (Bender *et al.*, 2006; Kraytsberg *et al.*, 2006; Hegde *et al.*, 2012), whereas post-mitotic tissues are also more likely to accumulate rearrangement mutations exponentially with age (Coskun *et al.*, 2003).

Heteroplasmy regulation is still not fully understood, especially during early development, or why the presentation of mitochondrial disease symptoms appear in individuals at different ages to a varying degree of severity, even within families (Crimmins *et al.*, 1993; Chinnery *et al.*, 1997; Nesbitt *et al.*, 2013). For example a mother could be mildly affected whereas her daughter/son could be severely affected (Ma *et al.*, 2013). One suggestion why heteroplasmy remains during evolution is because neither wildtype or mutant mtDNA is capable of fully supporting cell viability (Gitschlag *et al.*, 2016).

Clark and colleagues, and Taylor and colleagues suggest certain mtDNA mutants are ‘selfish’ and harbour in species, despite being harmful to cells, as they are co-transmitted with wildtype mtDNA, giving rise to persistent heteroplasmy (Taylor *et al.*, 2002; Clark *et al.*, 2012). To support this theory, Payne and colleagues use deep-sequencing to confirm heteroplasmy is transmitted in low levels universally in both somatic and germline cells in healthy humans (Payne *et al.*, 2013). Somatic mutations accumulate at different rates depending on the tissue type (Wallace, 1999; De Benedictis *et al.*, 2000), and clonal expansion of these errors gives rise to varying degrees of neurological and age-related diseases.

1.6.3 Mitochondrial DNA Copy Number between Tissue Types

Every cell and tissue type has their own metabolic requirements and limitations meaning mtDNA levels are not static and can fluctuate, especially during development (Michaels *et al.*, 1982). In the early stages of oogenesis, mtDNA levels escalate severely before significantly decreasing before implantation and gastrulation (Shoubridge and Wai, 2007; Kelly *et al.*, 2012). The decrease in mtDNA enables undifferentiated cells to establish their mtDNA CN threshold before differentiating into

their respective cells with specific OXPHOS requirements. Lee and colleagues suggest this is regulated through methylation profiles and use DNA immunoprecipitation techniques to demonstrate that *POLG-A* (which directly modulates mtDNA CN) is highly methylated in pluripotent cells and correlates to low mtDNA CN levels, whereas terminally differentiated cells have much higher mtDNA CN but lower levels of methylation at *POLG-A* (Lee *et al.*, 2015). The cell type and function normally dictates the mtDNA levels due to metabolic demand and has also been suggested to be controlled by mitochondrial-targeted nucleases (Reddy *et al.*, 2015).

High energy demanding cells like muscle and neurons usually have high mtDNA CN, whereas low energy demanding cells like spleen and endothelial cells contain low mtDNA CN (Dickinson *et al.*, 2013; Malik and Czajka, 2013). Mammalian sperm cells contain very little mtDNA CN compared to oocytes. Sperm contain about 50-75 copies per cell versus thousands of copies per cell in oocytes (Michaels *et al.*, 1982; Wallace *et al.*, 1999), although these values have been reported to be higher depending on cell cycle stage (Harvey *et al.*, 2011). Leucocytes have a high turnover rate and a lower metabolic demand than neurons, so they tend to accumulate less mtDNA mutations and have consistent mtDNA CN (Pyle *et al.*, 2010). White blood cells also contain less mtDNA than skeletal muscle, heart or kidney, and mtDNA CN is regulated independently to cytochrome c or citrate synthase activity levels (Van den Bogert *et al.*, 1993). Embryonic stem cells are also highly proliferative yet contain low mtDNA CN, suggesting mtDNA CN is not solely a product of OXPHOS demand (Facucho-Oliveira *et al.*, 2007).

Varying mtDNA CN between tissue types is believed to be a result of differential expression of the three different isoforms of *PPAR*s which are coded by individual nuclear genes (Berger and Moller, 2002). Lehman and colleagues also suggest PGC1 partners like NRF1 and PPAR control levels of expression of mitochondrial maintenance components, and are responsible for mtDNA CN variability between cell types (Lehman *et al.*, 2000).

1.6.4 Importance of Regulating Mitochondrial DNA Copy Number

In addition to homoplastic and heteroplasmic genetic variability, mtDNA CN has been argued to be the most important determinant of cellular homeostasis, especially during development (Shoubridge and Wai, 2007; Van Blerkom, 2011). Varied

heteroplasmic segregation has been observed through the bottleneck effect (Bogenhagen *et al.*, 2008), and nucleoid organisation has been proposed to contribute to this, using dynamic fusion and fission to ensure nucleoids have access to a larger pool of replication machinery (Bogenhagen *et al.*, 2008). Albeit not fully understanding how mitochondria are selected for or assigned to daughter cells through the bottleneck effect, Jarou and colleagues demonstrate that mitochondria and mtDNA nucleoids in yeast segregate and distribute evenly during mitosis relative to the cellular cytoplasm (Jajoo *et al.*, 2016), suggesting mtDNA CN segregation may be due to chance.

Research still has many unanswered questions surrounding the bottleneck effect, clonal expansion, and disease development - all of which strongly rely on mtDNA CN regulation but are not fully understood. mtDNA CN has been investigated as a biomarker of disease in a number of tissues for several years – reviewed by (Malik and Czajka, 2013). While mtDNA CN differences are being reported in diseases such as cancer, T2DM, and HIV; mtDNA levels are inconsistent between studies, suggesting mtDNA CN may be an indicator of disease progression instead of being a disease biomarker.

Intergenic communication has been suggested to be similar to quorum sensing, like their bacterial ancestors would do to synchronise gene expression (Ng and Bassler, 2009). With mtDNA CN being shown to act as both a regulator and a protector of disease development and progression, it is important to identify how mtDNA CN is regulated intergenomically, especially in relation to disease.

1.7 Mitochondrial DNA Copy Number in Health and Disease

Alterations in mitochondrial function contribute to the development of many metabolic-related diseases like T2DM (Silva *et al.*, 2000) and cancer (Reznik *et al.*, 2016), and contribute to the development of complex mitochondrial diseases such as Leigh’s syndrome, Leber’s hereditary optic neuropathy (LHON), MERRF, and MELAS (Wallace et al., 1988; Santorelli et al., 1993). Additionally, new research is showing that mitochondrial defects contribute to autoimmune disorders (Mills *et al.*, 2017; Sadatomi *et al.*, 2017). As reviewed by Malik and Czajka, mtDNA CN has been measured in numerous metabolic diseases, across multiple tissue types, and mtDNA levels are inconsistent across studies (Malik and Czajka, 2013).

Mitochondrial biogenesis has been reported to be different between sexes (Khalifa *et al.*, 2017; Mauvais-Jarvis *et al.*, 2017) and increasing evidence is emerging to show that biogenesis efficiency and response to insult is more important to disease progression. Increased levels of mtDNA are being reported to offer a protective effect against developing disease. For example, Wang *et al.* (2012) report that survivors of lepromatous leprosy had a significantly higher mtDNA CN than those who did not recover (Wang *et al.*, 2012). A similar study on rats also showed neuro-protective mechanisms being induced as mtDNA CN increased, despite an increase in mtDNA deletions in dopaminergic neurons being present which were similar to levels observed in PD patients (Perier *et al.*, 2013a). Cancer patients have been reported to have increased mtDNA CN too whereas HIV patients have depleted mtDNA CN (Lan *et al.*, 2008).

1.7.1 Cancer

Cancer is a term used to describe a disease which involves abnormal and uncontrollable cell proliferation in a specific tissue which has the ability to metastasise and spread to other parts of the body. Over 200 types of cancer exist but breast, lung, prostate, and bowel cancer are the most common (NHS, 2017) and in the UK, more than one in three people will develop cancer in their lifetime.

Cancer cells require functional mitochondria to survive. This has been shown in a number of studies using p^0 cell lines (reviewed (Wallace, 2012)). The relationship mtDNA has with cancer development has also been extensively researched in current literature. In NCBI, a search consisting of ‘cancer and mtDNA’ returns 8345 results, of which 318 results cover ‘mtDNA copy number’.

Despite the number of studies, the literature continually reports differences in mtDNA CN, even in the same type of cancer (Malik and Czajka, 2013; Reznik *et al.*, 2016). Xia *et al.* suggest mtDNA CN changes in females with cancer because of their menstruation status, causing changes in expression of oestrogen receptors (Xia *et al.*, 2009), although the range in their results is high and there are no controls. Mi J. and colleagues meta-analysed literature results from 38 independent study groups and reported mtDNA CN was *“consistently overrepresented in cancer cases relative to controls in both genders”,* but again, their results do not support this as mtDNA CN was also reported to be lower in certain cancers compared to their controls, supporting that seen in Malik’s review (Mi *et al.*, 2015). This may be because many

different types of cancer studies were combined and subjects were very heterogeneous, where ethnic origin, sample type, and gender comparisons were not consistent.

In colorectal cancer alone, risk has been associated with mtDNA CN and mortality, and affected individuals have amplified aberrant somatic mtDNA mutant variants (Polyak *et al.*, 1998; Srinivasainagendra *et al.*, 2017), however most cancers are so heterogeneous and physiologically complex these associations are not representative across all cancer types. mtDNA CN changes have been suggested to be an indication of disease stage, but this is undetermined (Lee *et al.*, 2005).

1.7.2 Mitochondrial Depletion Syndromes

Similar to cancer, when mtDNA becomes unstable and depleted, it leads to disease (Lee *et al.*, 2005). Profound depletion of mtDNA CN also results in MDS causing severe myopathic, encephalomyopathic and hepatocerebral clinical symptoms (Alberio *et al.*, 2007), and can often result in organ failure or death (Wallace, 1999). MDS signs have also presented in amniotic fluid leading to early-infant death (Blake *et al.*, 1999).

Mitochondrial neurogastrointestinal encephalopathy (MNGIE) disease is a well characterised condition that often develops by the age of 20 and affects several parts of the body, especially the digestive tract and nervous system (Zeviani and Di Donato, 2004). MNGIE develops from a mutation in the *TYMP* gene which codes thymidine phosphorylase, which contributes to reduced dNTP pools, resulting in decreased mtDNA CN as mtDNA replication is disrupted (González-Vioque *et al.*, 2011; Vondráčková *et al.*, 2014). The importance of dNTP pools nucleotide turnover and their direct effect on mtDNA CN regulation have been studied in non-diseased models elsewhere too (Nikkanen *et al.*, 2016). A group in Barcelona originally published that treatment with dNTPs was not a suitable pharmacological approach to relieve MDS symptoms, however recently republished results appear to suggest otherwise (González-Vioque *et al.*, 2011; Cámara *et al.*, 2014).

As previously discussed, any Mendelian defects in autosomal nuclear genes affecting nuclear-mitochondrial communication affect the stability of mtDNA by disrupting mtDNA replication and maintenance, and result in decreased mtDNA CN and stressed mitochondria (Limongelli and Tiranti, 2002).

1.7.3 Parkinson's disease

PD is an age-related neurodegenerative disease which affects around 0.3% of the global population, and about 1% of people over 60 years of age (Nussbaum and Ellis, 2003; de Lau and Breteler, 2006). PD is the second most common neurodegenerative disease after AD and develops as a result of dopamine deficiency in the striatum from the loss of dopaminergic neurons in the substantia nigra pars compacta (Winklhofer and Haass, 2010).

mtDNA CN has been reported to be significantly lower in peripheral blood, substantia nigra pars compacta, and CSF in individuals with PD (Pyle *et al.*, 2015a; Pyle *et al.*, 2015b). Increased mtDNA CN on the other hand (and improved OXPHOS) were reported to serve neuroprotective effects in substantia nigra pars compacta dopaminergic neurons induced with PD-like insults (Perier *et al.*, 2013b), however this could be haplogroup specific. Whilst haplogroups have been seen to regulate the bioenergetics of mitochondria (Gomez-Duran *et al.*, 2010; Gómez-Durán *et al.*, 2012) no large-scale study has investigated the differences in mtDNA CN between haplogroups until now.

Stressed mitochondria are also reported to leak mtDNA into the cellular cytosol and stimulate expression of immunological-factors such as interferons and antiviral signalling-elements (West *et al.*, 2015). Recent studies have shown circulating, cell-free mtDNA (ccf-mtDNA) in the CSF of PD patients, and microglia-associated inflammation is a pathological marker of PD, and is considered a potential therapeutic target (Kannarkat *et al.*, 2013; Su and Federoff, 2014; Wang *et al.*, 2015).

1.7.4 Type 2 Diabetes Mellitus

T2DM is a long-term metabolic disorder where insufficient insulin is produced, resulting in elevated blood sugar levels (NHS, 2017). T2DM is one of the most prevalent, chronic diseases worldwide, estimated to affect 347 million people globally in 2011 (Danaei *et al.*, 2011). T2DM is managed with medication and lifestyle changes, and has been proven to be reversed when individuals follow a calorie restricted diet (Lim *et al.*, 2011).

Similar to cancer, T2DM is a very complex and heterogeneous disease, and both nDNA and mtDNA genes are associated with developing it (Brandon *et al.*, 2005; Flannick and Florez, 2016; Fuchsberger *et al.*, 2016). mtDNA CN has also been investigated in T2DM, and like cancer, mtDNA CN is both higher and lower in

affected individuals than controls (Malik and Czajka, 2013). Low mtDNA CN has been considered to contribute to the development of T2DM as it is related to insulin resistance, age of onset and complications, and was recently considered a relevant predictor of T2DM (Cho *et al.*, 2017), however their diseased group was very small (containing 103 individuals).

What's more is diabetes is often a symptom of mitochondrial-related diseases such as MELAS, and maternally inherited diabetes and deafness (MIDD), and develops as a result of accumulated mutant mtDNA with age (Maassen *et al.*, 2004) (Taylor and Turnbull, 2005; Jacobs *et al.*, 2006).

As insulin is known to activate a cascade of metabolic signals within eukaryotes, it is difficult to determine whether mtDNA CN is disease-causing or if it is a biomarker of disease and physiological stress enforced on cells.

1.7.5 Leber's Hereditary Optic Neuropathy

LHON is the most common mitochondrial disease, affecting 1 in 14000 men predominantly harbouring either the 3460G>A mutation in *MT-ND1*, 11778G>A mutation in *MT-ND4* or 14484T>C mutation in *MT-ND6*, all which code for subunits forming complex I of the OXPHOS system (Man *et al.*, 2003; Pello *et al.*, 2008).

LHON is a homoplasmic disease characterised by rapid onset of sudden bilateral blindness due to retinal ganglion cell loss, predominantly in the papillomacular bundle (Hudson *et al.*, 2007a; Russell and Turnbull, 2014). Visual loss is more likely to occur in haplogroup J individuals when harbouring the 11778G>A or 14484T>C mutations, or haplogroup type K individuals when harbouring the 3460G>A mutation. Contrariwise, haplogroup H has protective effects for those with the 11778G>A mutation (Hudson *et al.*, 2007b).

Only 40% males and 10% females who harbour one of the mutations develop LHON, showing incomplete penetrance. Also, approximately 12% of those LHON patients are heteroplasmic to both wild-type and mutant DNA, and heteroplasmy has been seen to reduce the risk of blindness (Man *et al.*, 2003). Hudson G. *et al.* showed there is no difference in LHON development irrespective of age or X-inactivation, so the X chromosome cannot be the sole cause of incomplete penetrance (Hudson *et al.*, 2007a). This in combination with Giordano and colleagues work suggesting there

are other contributing factors to disease penetrance, and this may involve mitochondrial biogenesis regulation.

Giordano and colleagues measured mtDNA CN and mass in patients affected by, or carriers of, LHON mutants, and reported that carriers had a significantly higher mtDNA CN than those of their affected relatives (Giordano *et al.*, 2014). This suggests that the excessive mitochondrial biogenesis may have protective effects to compensate for the pathogenic mtDNA mutants, and carriers must have cellular mechanisms regulating mitochondrial biogenesis more efficiently. This emphasises the need to better understand mtDNA CN and nuclear communication in regulating mitochondrial biogenesis.

Genetic regulation of mtDNA CN is not well understood across populations, and most genetic studies published are in a small number of individuals or cell lines, which is not representative of larger populations. Therefore we designed a study which was able to develop on these limitations.

1.8 Genome-wide Association Studies

Although being a significant component of this study, an exhaustive description of genome-wide association studies (GWAS) is beyond the scope of this thesis. Rather I present a brief description aimed at guiding the reader through some of the key points. Interested readers should refer here for further information (Burton *et al.*, 2007; McCarthy *et al.*, 2008; Wang *et al.*, 2010a; Barsh *et al.*, 2012).

GWAS are studies where large sets of genetic DNA markers (normally SNPs) are genotyped and examined in large populations of unrelated individuals, to identify common genetic variants which are associated with a specific phenotype. Published GWAS have rapidly increased since genotyping microchips were introduced in the early noughties (Figure 1.13A), and from comparing diseased and control cohorts, hundreds of thousands of SNPs have been identified to be associated with disease across most human genomes (Figure 1.13B) (The Wellcome Trust Case Control Consortium, 2007; Hindorff LA. *et al.*, 2017; MacArthur *et al.*, 2017). In combination with the international HapMap project, which introduces human chromosome variability across different populations, it has made it easier to investigate genomic differences by continent too (Belmont *et al.*, 2005).

GWAS are becoming increasingly popular as the cost of whole-genome sequencing has decreased and the availability of genotyping data across large populations has increased. As a result, GWAS provide a powerful platform to invest complex traits. From the thousands of published GWAS papers, many of which have identified genetic variants, both rare and common, causing complex traits and disease (Khor *et al.*, 2011; Nalls *et al.*, 2014; Schizophrenia Working Group of the Psychiatric Genomics, 2014; Coffee *et al.*, 2015; Farh *et al.*, 2015; Shungin *et al.*, 2015; Fuchsberger *et al.*, 2016; Long *et al.*, 2017).

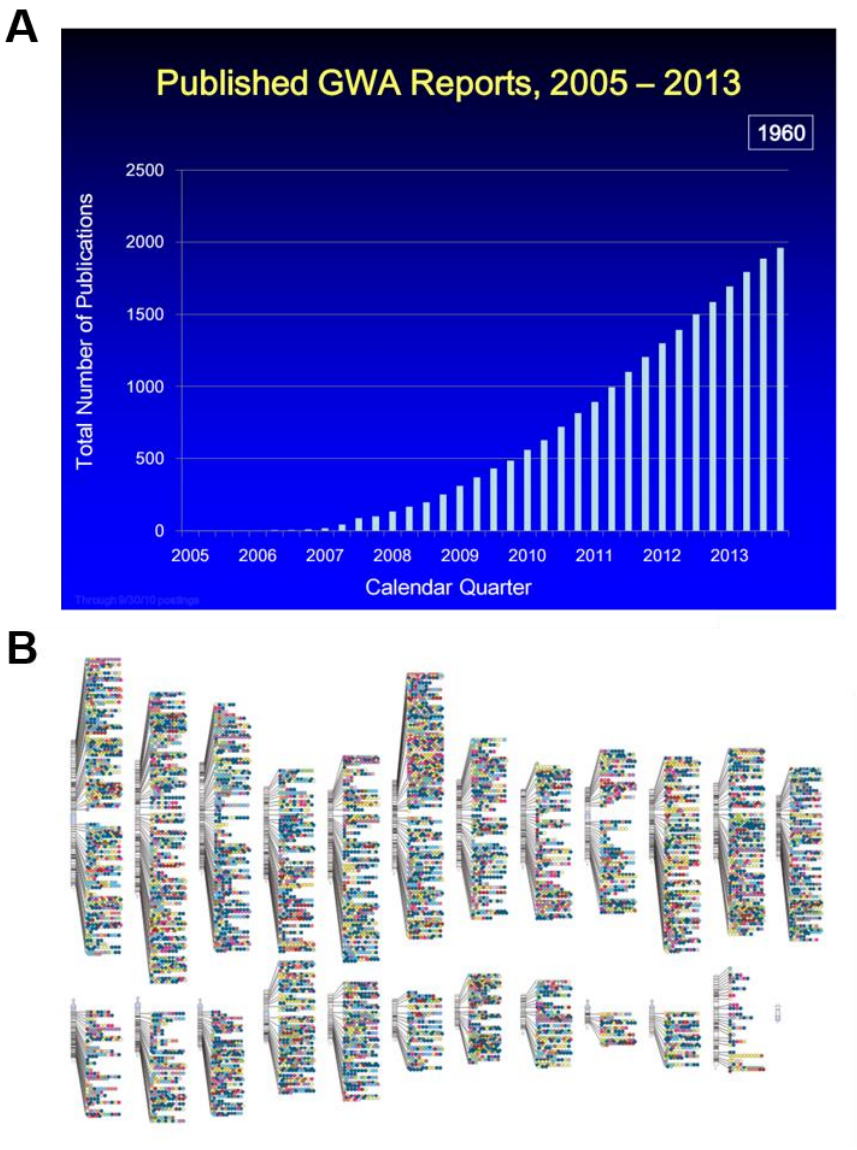


Figure 1.13: [A] Number of published genome-wide association studies from 2005 to 2013 and [B] the EMBL GWAS catalogue showing all published SNP-trait associations (with p-values ≤5.0 x 10^{-8}) to their respective chromosome location. Each linear grey box corresponds to a single chromosome. Coloured spots represent individual SNPs and their location. Images from (Hindorff LA. et al., 2017; MacArthur et al., 2017).

Despite the power and popularity of GWAS, they still contain limitations. To start, most GWAS use data generated from DNA extracted from blood. This provides limitations when researching diseases which are tissue specific. Additionally, common, yet very heterogeneous traits like height, cancer, and T2DM are still not fully understood, despite the number of GWAS performed (Aulchenko *et al.*, 2009; Marouli *et al.*, 2017). This is believed to be a result of traits not being 100% heritable. For example T2DM is only believed to be 10% heritable (Billings and Florez, 2010; Marouli *et al.*, 2017). Further to this, GWAS are used to study common genetic variants across populations, and are limited to identify rare variants (Figure 1.14), and reproducibility of GWAS results are very few compared to the number of studies

published. Furthermore they only identify genetic loci which are associated to a phenotype, they do not explain the full biological picture as to how the phenotype presents. Thus GWAS are only a preliminary indicator of risk loci and suggest where to start further biological investigations (McCarthy *et al.*, 2008; Ward and Kellis, 2012).

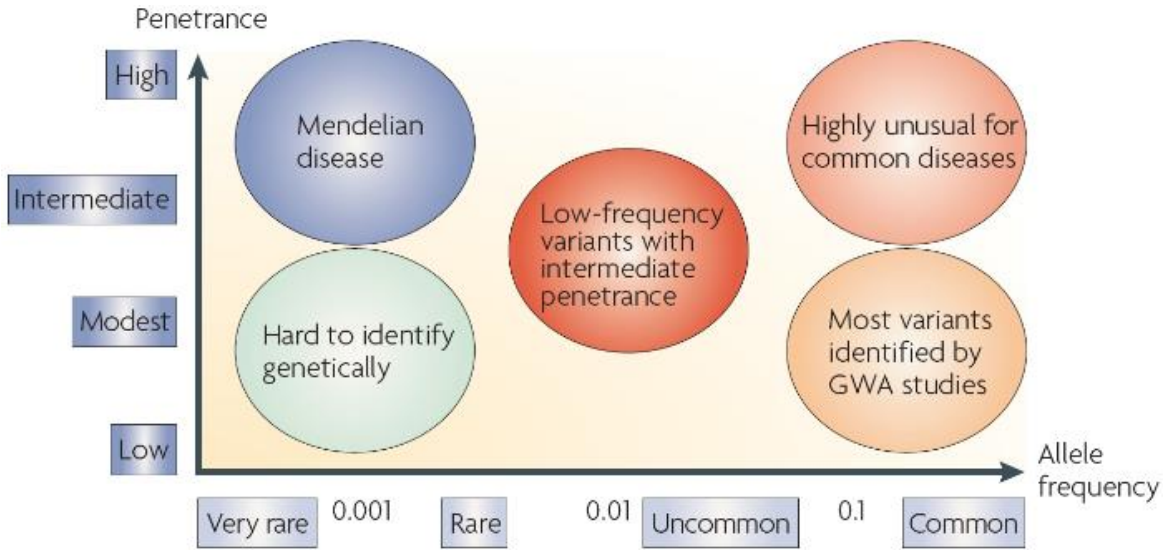


Figure 1.14: Image summarising the challenges behind detecting selection bias. Image produced by (McCarthy et al., 2008).

GWAS, like most research methods, have their limitations, however their power to identify genetic variability and associations to phenotypes across populations is also very advantageous. As literature shows, regulation of mtDNA CN is still not fully understood, but is an important system to understand, especially in preventing, treating and understanding disease. The literature is flooded with biological evidence explaining the complex physiological networks known to cause mtDNA CN changes, no definitive study has identified key genes involved in intergenomic communication between nDNA and mtDNA in regulating mtDNA CN.

As mtDNA CN is an easily measured phenotype that has been reported to be regulated by common genetic variants, we chose to investigate mtDNA CN by performing the largest ever GWAS using actual measures of mtDNA CN as the phenotype. Using both mtDNA and nDNA SNP genotyping data in thousands of healthy, unrelated, European individuals we have investigated mtDNA CN variability across populations and have identified potential nuclear and mitochondrial candidate genes involved in mtDNA CN regulation. In addition to this we have investigated the accuracy of current methods to measure mtDNA CN and have identified technical alterations that need to be considered for future research surrounding mtDNA CN.

1.9 Hypothesis

mtDNA CN is not one trait because mtDNA CN varies between cells, tissues, and organs depending on cellular metabolic demand and mitochondrial quality. It is therefore rational to hypothesise that mtDNA CN is under significant genetic control and is influenced by biological factors. In this study, mtDNA CN will be quantified from DNA extracted from white blood cells (WBC) in peripheral blood, and this single phenotype will be representative of mtDNA CN regulation in a preliminary study to identify genetic and biological factors associated with mtDNA CN changes. By performing genetic association tests, using population variants in large well-characterised cohorts, it will be possible to determine nuclear and mitochondrial DNA variants modulating mtDNA CN. Additionally, the effects biological factors such as age, gender, blood composition, oestrogen, and mtDNA deletions have on mtDNA CN levels will be determined.

1.10 Aims

1. Compare and contrast existing mtDNA real-time PCR quantification methods, assessing the validity of existing methods in multiple large population cohorts
2. Use real-time PCR (informed by (1)) to investigate the degree of mtDNA CN variation in DNA samples extracted from WBC of unrelated, healthy individuals in multiple large, European populations.
3. Compare mtDNA CN in large populations to demographic data such as age and gender, and phenotypic measurements such as blood composition to identify significant confounding factors modulating changes.
4. Conduct a large-scale GWAS analysis of mtDNA CN in multiple large population cohorts, comparing nDNA and mtDNA genotypes to mtDNA CN as a quantitative trait locus (QTL).
5. Perform an imputation analysis of (4) to increase our genetic resolution and power, and compare established imputation methods.
6. Perform a replication of (4) in a second large, unrelated cohort - utilising existing genotype data available through the Avon Longitudinal Study of Parents and Children (ALSPAC) through collaboration with Dr Santiago Rodriguez (University of Bristol, UK) and further increase our statistical power.
7. Identify key genetic loci associated with mtDNA CN changes in WBC to investigate in future functional studies to help further understand the biological

mechanisms managing mitochondrial maintenance, and contributing to disease development.

Secondary Aims

8. Explore the degree of protection oestrogen has on mitochondrial biogenesis by comparing mtDNA CN in pre- and post-menopausal European women.
9. Examine the quality of mtDNA in large populations of healthy individuals, and investigate the prevalence of mitochondrial deletions and their effect on mtDNA CN.

Chapter 2 Materials and Methods

2.1 Cohorts

2.1.1 Cohort One - United Kingdom Blood Service Cohort

In 2005, the Wellcome Trust Case Control Consortium (WTCCC) set up a UK national repository in collaboration with the UK Blood Service (hereafter referred to as the UKBS cohort) (compromised of NHS Blood Transplant in England, Scottish National Blood Transfusion Service in Scotland, and Welsh Blood Service in Wales) where 3622 anonymised samples of DNA and mononuclear cells were collected. In order to donate blood to these services, a series of strict regulations have to be met; where current health and medications being taken, travel outside of the UK, tattoos and piercings, and pregnancies are all limitations (NHS Blood and Transplant., 2017). These strict restrictions ensure that the blood samples provided are 'normal', and representative of a typical, healthy individual in the population.

2.1.2 Cohort Two - In house 'Cohort Two'

As part of ongoing work in our laboratory, I had access to 2000 DNA samples from 1000 mother and child pairs which had been stored at -80ºC from previous work.

2.1.3 Cohort Three – Avon Longitudinal Study of Parents and Children Cohort

A collaboration was organised between Dr Santi Rodriguez's group, in the School of Social and Community Medicine at University of Bristol, UK, and Newcastle University to use the Avon Longitudinal Study of Parents and Children (hereafter referred to as ALSPAC) cohort as a replication cohort for the work in this study. ALSPAC is a prospective cohort of mothers and their children, and was established to understand how genetics and environment influence health and development in parents and children. 5461 ALSPAC mothers were used in this study.

2.2 DNA Sample Source

2.2.1 Control Cohort One - UKBS

Blood was collected from healthy, unrelated donors across 12 Great British health regions between September 2005 - February 2006 by UK National Blood Services (UKBS). Participant's ages ranged between 17-69 years, and were recruited based on sex and geographical region. Informed consent was obtained from all participants in accordance with protocols approved by the Peterborough & Fenland Local Research Ethics Committee in September 2005 (approval 05/Q0106/74) (Burton *et al.*, 2007).

2.2.2 Cohort Two - In house 'Cohort Two'

DNA was sourced from 1000 mother and child pairs living in the former county of Avon, UK. Age and gender of samples was not available however mothers could be assumed to be female.

2.2.3 Cohort Three – ALSPAC

From 1st April 1991 to 31st December 1992, 13,761 women were recruited during pregnancy (aged between 16-43 years), who were living in the former county of Avon, UK (Boyd et al., 2013; Fraser et al., 2013). For the duration of the study, these women had to complete up to 20 detailed questionnaires, and had data extracted from their medical records. Full details of all the data available is accessible online: http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

2.3 Generation of Blood Count Data

2.3.1 Cohort One - UKBS

Full blood count data (FBC) was generated from ethylenediaminetetraacetic acid (EDTA) anticoagulated samples of blood drawn from the pouches of the donation collections. Samples were transferred at room temperature to a single testing centre in Cambridge, and FBC measurements were generated 16-24 hours after phlebotomy using a Beckman-Coulter GenS (Gieger *et al.*, 2011).

The Beckman-Coulter GenS uses fluorescence-activated cell sorting technology to separate and count each cell type. Individual cells are forced through an aperture tube by a vacuum which is suspended in some conductive liquid. As each blood cell moves through the aperture tube and liquid, the blood cell insulates the current and causes an increase in electrical resistance which generates a pulse. Each pulse correlates to one cell, and the size of the resistance recorded is proportional to the cell volume (Figure 2.1). Following this, a series of low and high frequency currents measure internal and external cell volume respectively, and then light and the addition of cell-specific staining dyes identify individual white blood cells (WBC) (Beckman Coulter Inc., 2011a). Total WBC count was reported in thousands of cells per ml, and sub-type specific cell counts were calculated by multiplying the proportion of the WBC count, comprised by each cell type, by the total WBC measure (Nalls *et al.*, 2011). Blood data consisted of raw cell counts and whole blood cellular

proportions for basophils, eosinophils, lymphocytes, monocytes, and neutrophils, as well as platelet (PLT) count, white blood cell (WBC) count, red blood cell (RBC) count, haematocrit (HCT), haemoglobin (HB), mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), and mean corpuscular hemoglobin concentration (MCHC) (Table 2.1).

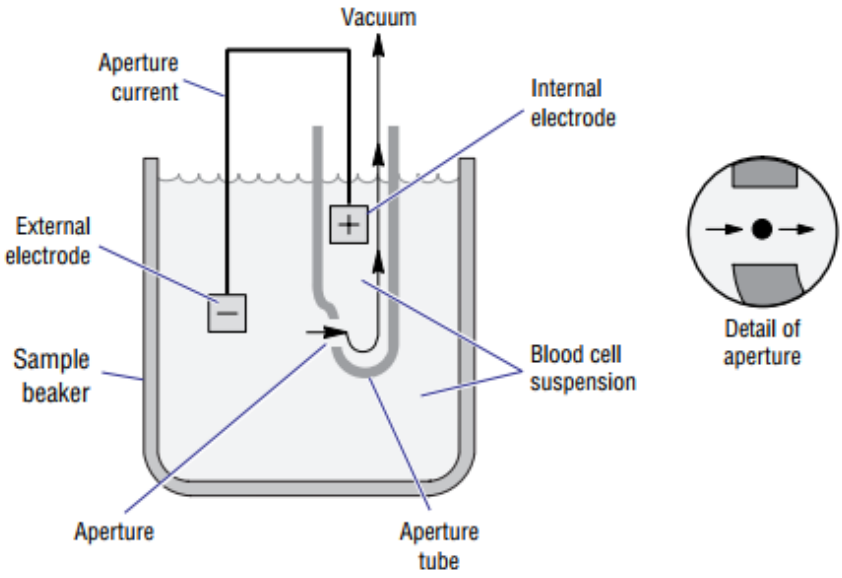


Figure 2.1: Diagram of how the Beckman-Coulter GenS counts and sizes individual blood cells. Image taken from (Beckman Coulter Inc., 2011a).

Through collaboration with Professor Willem H Ouwehand at University of Cambridge, FBC data was provided to Newcastle University for 2617 UKBS samples. FBC data is numeric; however absolute basophil, and absolute eosinophil counts included a verbose output stating “less than 0.01”. These data (n=2337, and n=180 respectively) had a value of zero assigned because I was unable to determine the exact values. Absolute basophil count was then considered a binary trait, because values were either zero or 0.1. Details of the FBC abbreviations, and their respective units were provided in the Beckman Coulter manual, and are detailed in Table 2.1 (Beckman Coulter Inc., 2011b).

Blood Shorthand	Verbose Blood	Units	Number of Individuals With Data	"Less than 0.01"	Missing	Total
PLT	Platelet Count	10^3 cells/µL	2603	0	488	3091
RBC	Red Blood Cell Count	10^6 cells/ µL^-1	2617	0	474	3091
HB	Haemoglobin	g/dL	2617	0	474	3091
HCT	Haemocrit	%	2617	0	474	3091
MCH	Mean Corpuscular Haemoglobin	pg	2617	0	474	3091
MCHC	Mean Corpuscular Haemoglobin Concentration	g/dL	2617	0	474	3091
MCV	Mean Corpuscular Volume	10^-15 L	2617	0	474	3091
WBC	White Blood Cell Count	10^3 cells/ µL^-1	2617	0	474	3091
BA	Absolute Basophils	10^3 cells/µL	280	2337	474	3091
BP	Basophil Proportion	%	2617	0	474	3091
EA	Absolute Eosinophils	10^3 cells/µL	2437	180	474	3091
EP	Eosinophil Proportion	%	2617	0	474	3091
LA	Absolute Lymphocytes	10^3 cells/µL	2617	0	474	3091
LP	Lymphocyte Proportion	%	2617	0	474	3091
MA	Absolute Monocytes	10^3 cells/µL	2617	0	474	3091
MP	Monocyte Proportion	%	2617	0	474	3091
NA	Absolute Neutrophils	10^3 cells/µL	2617	0	474	3091
NP	Neutrophil Proportion	%	2617	0	474	3091

Table 2.1: Details of full blood count measurements generated by a Beckman‐Coulter GenS, and the number of UKBS samples with blood count data. Data was provided by Willem Ouwehand at University of Cambridge.

2.3.2 Cohort Two - In house 'Cohort Two'

Blood count data was not available for cohort two samples.

2.3.3 Cohort Three – ALSPAC

The Accessible Resource for Integrated Epigenomics subset (ARIES) is a sub-study of ALSPAC including 1018 mother-child pairs. Blood cell proportions were estimated from DNA methylation data which was derived from maternal antenatal samples as part of the ARIES study (Relton *et al.*, 2015). Samples were assayed on the Illumina Infinium HumanMethylation450 BeadChip (450K) array (Illumina, San Diego, California, USA), and cell count composition was determined from this estimated data using the Houseman method (Houseman *et al.*, 2012). Data was processed using the 'estimateCellCounts' function of the R Bioconductor package 'minfi' (Jaffe and Irizarry, 2014; Richmond *et al.*, 2015), and 546 ALSPAC mothers had estimated cell count data generated.

2.4 DNA Extraction and Preparation

2.4.1 Cohort One - UKBS

Following blood sample collections, DNA extractions were performed by the respective collection centre. Blood was separated by density centrifugation, and white blood cells were retained to perform phenol DNA extractions. Following extraction, samples were provided to the Wellcome Trust Sanger Institute where DNA quality was assessed on a 0.75% agarose gel, and was quantified using the PicoGreen® method to ensure heavily degraded samples were eliminated. DNA was

eluted in 10 mM Tris and 100 μM EDTA and normalised to 200 ng/μl per well. Full details are provided here (Burton *et al.*, 2007).

2.4.2 Cohort Two - In house 'Cohort Two'

1889 DNA samples (89% maternal and 100% child) were extracted from blood, and 111 (11.1%) maternal samples were extracted from buccal samples. DNA extraction and preparation details were not available at the time of writing (Table 2.2).

2.4.3 Cohort Three – ALSPAC

DNA was extracted from blood using a phenol-chloroform method (Fraser *et al.*, 2013). 5461 maternal samples were used in this study, of which 2056 samples (37.6%) had been extracted from whole blood and 3405 samples (62.4%) had been extracted from white cell pellets (Table 2.2).

2.5 Nuclear DNA SNP Genotyping

2.5.1 Cohort One - UKBS

UKBS DNA samples with a concentration of ≥50 ng/μl^{-1} which showed limited or no degradation were initially genotyped with a MassEXTEND™ and/or iPlex® assay which arrays up to 38 SNPs to experimentally confirm the sex of each sample. DNA samples that were initially arrayed successfully were sent for whole genome SNP genotyping. As part of the WTCCC2, SNP genotyping was performed on the Illumina 1.2M Duo array (Illumina, San Diego, California, USA), and were aligned to the NCBI36-hg18 human genome build (Wellcome Trust Sanger Institute, 2008). 1,152,686 variants were successfully genotyped for 2735 individuals of European-ancestry and was then provided to Newcastle University to be analysed.

2.5.2 Cohort Two - In house 'Cohort Two'

Cohort two samples were not genotyped for nuclear DNA SNPs.

2.5.3 Cohort Three – ALSPAC

All ALSPAC maternal DNA samples were genotyped on the Illumina Human660W-Quad array (Illumina, San Diego, California, USA) at the Centre Nationale du Génotypage. Genotypes were called using Illumina GenomeStudio® and were aligned to the NCBI36-hg18 human genome build. 526,688 variants were successfully genotyped. This data was then analysed at University of Bristol.

Cohort #	Cohort Name	N	Sample Type	Male: Female Ratio	Age Range (years)	DNA Source	Cell Type	DNA Isolation Method	Results Chapters
1	UKBS	3091	Control, unrelated individuals	50:50	17-69	Blood	WBC (100%)	Phenol chloroform	All
2	In house 'Cohort Two'	2000	Control, mother-child pairs	Unknown	Unknown	Blood (94.4%) & Buccal (5.6%)*	Unknown	Unknown	1 and 3
3	ALSPAC	5461	Control, mothers only	0:100	16-43	Blood	WB (37.6%) & WBC (62.4%)	Phenol chloroform	2 and 3

*Table 2.2: Summary of cohorts used in this study and their demographic details including ratio of males to females, age range (years), DNA source, the cell type DNA was extracted from, the DNA isolation method used, and the results chapters where each respective cohort was used for data analysis. UKBS: United Kingdom Blood Service Cohort. ALSPAC: Avon Longitudinal Study of Parents and Children Cohort. WB: Whole Blood. WBC: White Blood Cells. * indicates samples were extracted from blood samples but blood cell type was unknown and buccal samples were excluded for all analysis.*

2.6 Mitochondrial DNA Genotyping and Sequencing

2.6.1 Cohort One – UKBS

UKBS samples were genotyped as part of the WTCCC2 (Wellcome Trust Sanger Institute, 2008) using the Illumina 1.2M Duo SNP array (Illumina, San Diego, California, USA) which contains 138 mitochondrial SNPs (mtSNPs) (Hudson *et al.*, 2014). 132 mtSNPs were successfully genotyped (>99.0%) for 2735 individuals. This data was then provided to Newcastle University to be analysed.

2.6.2 Cohort Two - In house 'Cohort Two'

Cohort two samples were not genotyped, and instead were outsourced to Source BioScience™ in Nottingham, UK, where mtDNA was sequenced using mitochondrial-targeted Illumina next generation DNA sequencing. Library preparation was performed using PCR amplicon sequencing using a Fluidigm Access Array™ (Fluidigm, UK). Unique sequence-specific tags and sample-specific bar codes were added to each DNA sample and were pooled together and run on a single sequencing experiment. Using Access Array Integrated Fluidic Circuits™ (Fluidigm, UK), approximately 100 amplicons were designed (150 to 200 bp) to cover the entire mtDNA. 48 PCR reactions were performed on each sample. Following this, PCR products were purified and quantified before sequencing by an Agilent 2100 Bioanalyzer™ (Agilent, UK). PCR products were then pooled in equal volumes to create one PCR product library which were then purified by AMPure XP beads (Beckman Coulter Life Sciences, Buckinghamshire, UK). PicoGreen fluorimetry was used to quantify the PCR product library prior to loading for DNA sequencing. DNA sequencing was finally performed on the Illumina HiSeq 2000™ array (Illumina, San Diego, California, USA) using 100 bp paired-end reads, generating 150 million clusters.

1627 samples (81.35%) had their mtDNA successfully sequenced. Raw data files were then provided to Newcastle University and were subjected to the same bioinformatic analysis. Raw FASTQ reads were filtered using FASTQ (version 0.11.2) and were aligned to both the rCRS (NM_012920.1) and the human reference genome (hg19) using BWA (version 0.7.12). Duplicate reads were then removed using Picard (version 1.130). Subsequent variant calling was performed using VarScan (version 2.3.7). Coverage of the mtDNA was generated for all samples and samples <99% coverage were removed from further analysis. FASTQ files generated

using the 180-amplicon enrichment strategy (Source BioScience™ Nottingham, UK) were subjected to primer trimming using cutadapt (version 1.14). Exact primer sequences were then trimmed from the 5' end of each FASTQ read and subsequent trimmed FASTQ files were subjected to the same bioinformatic pipeline as above. This data was then used to assign mitochondrial haplogroups (section 5.3.2.2).

2.6.3 Cohort Three – ALSPAC

ALSPAC maternal samples were not genotyped directly for mtSNPs. Instead, 9912 ALSPAC children (male and female) were genotyped on the Illumina HumanHap550-Quad array, by 23andme, the Wellcome Trust Sanger Institute (Cambridge, UK), and the Laboratory Corporation of America (Burlington, North Carolina, USA), and were called for using the Illumina GenomeStudio®. The Illumina HumanHap550-Quad genotyping array contained 7554 custom mitochondrial probes, targeting 2824 unique mtDNA positions.

Using PLINK (version 1.07) software (Purcell *et al.*, 2007), individuals were excluded from further analysis for having incorrect gender assignments, minimal or excessive heterozygosity (<0.320 and > 0.345 for the Sanger data and <0.310 and >0.330 for the LabCorp data), disproportionate levels of individual missingness (>3%), evidence of cryptic relatedness (>10% IBD), and being from non-European ancestry (as detected by a multidimensional scaling analysis seeded with HapMap 2 individuals). EIGENSTRAT analysis revealed no additional obvious population stratification and genome-wide analyses with other phenotypes indicate a low lambda) (Price *et al.*, 2006). 8365 unrelated individuals passed QC analysis.

Prior to QC, all heterozygous genotype calls (i.e. heteroplasmy) were set to missing using PLINK (Purcell *et al.*, 2007). Genotype calls obtained from each probe were compared to the human mitochondrial database of non-pathological mitochondrial sequence variants (http://www.hmtdb.uniba.it:8080/hmdb/) to ensure that known allelic variants were being called. Probes were excluded in cases where genotype calls were not represented in the revised Cambridge Reference Sequence (rCRS) reference or one of the known allelic variants. Probes with an overall call rate of <95% were excluded prior to analysis. The genotyping concordance of the remaining probes was investigated by comparing the genotype calls in 445 replicate samples. With the exception of probe failure (i.e. missing data), a 100% genotyping concordance rate was obtained for each probe. All probes with a failure rate of >5%

in the replicate sample were further excluded. A total of 1062 probes passed QC for the batch that was genotyped by Laboratory Corporation of America (n=7590), whilst 629 probes passed QC for the batch genotyped by the Sanger Institute (n=775). 308 of these probes overlapped with each other across the two arrays. Where there were several probes genotyping the same mtDNA variant, we used a consensus of the probes, and removed all duplicates. This resulted in 105 unique variants, all with MAF >0.01 (29 with MAF >0.05). mtDNA genotyping QC analysis was conducted by Dr Mesut Erzurumluoglu as part of his PhD with Dr Santi Rodriguez in the School of Social and Community Medicine at University of Bristol.

As mtDNA is inherited maternally (Hutchison *et al.*, 1974; Giles *et al.*, 1980), ALSPAC children mtSNP genotypes were assigned to their respective ALSPAC mothers. 3405 mothers were assigned mtSNP genotyping data.

2.7 DNA Samples in Newcastle

2.7.1 Cohort One - UKBS

3091 genomic DNA samples were provided to Newcastle University, in two panels, from the University of Cambridge, and NHS Blood and Transplant UK, as part of the WTCCC2 control group (Wellcome Trust Sanger Institute, 2008). Upon arrival, samples were diluted to 2 ng/µl in nuclease-free water in 96 PCR well plates (translucent shell) (BioRad, Hertfordshire, UK) and were covered with aluminium- foil sealing tape (StarLab, Milton Keynes, UK), and then were stored at -20°C until use. All patient identities were anonymised.

2.7.2 Cohort Two - In house ‘Cohort Two’

Before analysis, all 2000 samples were defrosted from -80°C and diluted to 5 ng/µl in nuclease-free water in 384-well PCR plates (clear shell, white well) (BioRad, Hertfordshire, UK) and were covered with aluminium- foil sealing tape (StarLab, Milton Keynes, UK), and then were stored at -20°C until use.

2.7.3 Cohort Three – ALSPAC

All ALSPAC DNA samples were handled and analysed at University of Bristol. No ALSPAC DNA samples were analysed at Newcastle University, except in the quantitative real-time PCR (qPCR) assay validation test detailed in section 4.4.3.1.

2.8 Measuring Mitochondrial DNA Copy Number using Quantitative PCR

2.8.1 Cohort One - UKBS

In order to perform a QTL GWAS, using mtDNA CN as the phenotype, mtDNA CN was measured in all UKBS samples using qPCR. Beta-2-microglobulin (*B2M*) was used as the nuclear reference and housekeeping gene to correct for the number of cells in each extract, and mitochondrially encoded NADH-ubiquinone oxidoreductase core subunits 1 and 4 genes (*MT-ND1* and *MT-ND4* respectively) and mitochondrially encoded 16S ribosomal RNA (*MT-RNR2*) acted as references for the number of mtDNA molecules. *MT-ND1* and *MT-ND4* were multiplexed with *B2M* in a triplex reaction for all qPCR assays, except where an 'alternative' qPCR assay is detailed in section 5.3.6 and *MT-RNR2*, *MT-ND4*, and *B2M* were multiplex together (Table 2.8).

As *MT-ND4* is diametrically opposed to *MT-ND1* and *MT-RNR2* in mtDNA, we could calculate mtDNA CN relative to each gene. This acted as a quality control step for calculating mtDNA CN and helped monitor the quality of the mtDNA, and the efficiency of the run, and enabled the calculation of mitochondrial deletion levels (He *et al.*, 2002; Krishnan *et al.*, 2007; Grady *et al.*, 2014a; Rygiel *et al.*, 2015).

2.8.1.1 Polymerase Chain-Reaction (PCR) and Control Template Preparation

To generate standard curves, templates of each gene of interest were generated to validate amplification specificity and assay efficiency. Primers were designed to *B2M* (Genbank accession ID: NM_004048), *MT-ND1, MT-ND4,* and *MT-RNR2* (Genbank accession IDs: *MT-ND1*, *MT-ND4,* and *MT-RNR2* NC_012920). Templates were generated from control DNA which was extracted from whole blood from healthy, anonymous donors using the Illustra® DNA extraction kit (GE Healthcare Life Sciences, Buckinghamshire, UK). Control DNA was amplified by PCR using gene-specific primers (Integrated DNA Technologies, Leuven, Belgium) in Table 2.3.

Gene	Amplicon Size (bp)	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
B2M	1092	CGCAATCTCCAGTGACAGAA	GCAGAATAGGCTGCTGTTCC
MT-ND1	1040	CAGCCGCTATTAAAGGTTCG	AGAGTGCGTCATATGTTGTTC
MT-ND4	1072	ATCGCTCACACCTCATATCC	TAGGTCTGTTTGTCGTAGGC
MT-RNR2	1026	CAGTGACACATGTTTAACGGC	GGAGGGGGGTTCATAGTAG

Table 2.3: A list of forward and reverse primer 5'-3' sequences used to generate standard template from control DNA using PCR. PCR templates amplified are nuclear gene beta-2-microglobulin (B2M); (Genbank accession ID: NM_004048) and mitochondrially encoded NADH-ubiquinone oxidoreductase core subunits 1 and 4 genes (MT-ND1 and MT-ND4 respectively) and mitochondrially encoded 16S ribosomal RNA (MT-RNR2) (Genbank accession IDs: MT-ND1, MT-ND4, and MT-RNR2 NC_012920). All primers were provided by Integrated DNA Technologies in Leuven, Belgium.

Reaction conditions for PCR master mix were: 0.4 µM of each forward and reverse primer, 1x MyTaq™ reaction buffer (contains MyTaq™ buffer, dNTPs, $MgCl_2$, enhancers and stabilizers), and one unit MyTaq™ DNA polymerase (Bioline, London, UK). All reactions used Ambion® nuclease free water (Thermo Fisher Scientific, Loughborough, UK) and were set up on ice. Approximately 50 ng/µl DNA was loaded into the PCR mastermix, which was then amplified on an Applied Biosystems® Veriti® 96 well thermal cycler in Applied Biosystems® MicroAmp® optical 96-well PCR plates (both Life Technologies, Paisley, UK). The PCR cycling conditions were; initial activation at 95°C for one minute, then 30 cycles of denaturation at 95°C for 15 seconds, annealing at 61°C for 15 seconds, and extension at 72°C for 10 seconds, followed by a final extension step at 72°C for 10 minutes. Negative controls were used to eliminate any potential contamination.

2.8.1.2 Agarose Gel Electrophoresis

Following the amplification of standard DNA templates, agarose gel electrophoresis was used to visually inspect and isolate the template PCR products. Multiple PCRs (typically n=10) for each template product were pooled and combined with 20% v/v orange G loading buffer (Orange G powder, 50% v/v glycerol (both Sigma Aldrich Company Ltd, Dorset, UK) and 50% v/v water). A negative PCR reaction was mixed 50% v/v with orange G loading buffer, as was a 1Kb DNA ladder (Promega, WI, USA). Positive template PCR products were loaded into a single large well on a 1% agarose gel (1% w/v agarose (Bioline, London, UK) in 1x tris-acetate-EDTA (TAE) buffer (5 Prime GmbH, Hilden, Germany), and 0.4 mg/µl UltraPure™ ethidium bromide (Invitrogen, Paisley, UK)). Ladder and negative products were loaded in

separate wells. Agarose gels were electrophoresed at 65 V for 90 minutes in 1x TAE buffer.

2.8.1.3 Template PCR Purification and Quantification

Agarose gels were imaged and analysed using the UVP GelDoc-It™ imaging system and UVP Launch Vision Works LS software (UVP, Cambridge, UK). The pooled template PCR products were first sized against the molecular weight DNA ladder to check that the template PCR products were specific, and of the correct size. Template PCR products were then isolated from the gel using the UVP Syngene Ultraviolet Gel Tray (UVP, Cambridge, UK). The gel extract was purified using QIAquick Gel Extraction Kit; micro-centrifugation protocol – according to manufacturer's instructions (QIAGEN, Manchester, UK). Extract concentrations were calculated by a Nanodrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific, Loughborough, UK) using Nanodrop 2000 software, measuring concentration/µl and 260/280 nm wavelength ratio with a baseline correction of 340 nm. Template PCR product purity was assessed using both the 260/280 ratio, where measurements between 1.8 and 2.0 were accepted, and the 260/230 ratio, where measurements between 2.0 and 2.2 were accepted. The DNA concentration was used to calculate DNA copy number of the template with Equation 2.1. Template DNA was stored at -20°C until use.

$$Copy\ number = [C\ \div (L\ x\ 2\ x\ 330)]\ x\ A$$

Equation 2.1: Formula used to calculate DNA copy number (copies/µl) once template DNA was extracted and concentration was determined by Nanodrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific, Loughborough, UK) in ng/µl. C is DNA concentration in nanolitres (10-9), L is amplicon length in base pairs and A is Avogadro's constant (6.023 x 10^23).

2.8.1.4 Quantitative Real-Time PCR

To quantify mtDNA CN, qPCR was used. Quantifying *MT-ND1*, *MT-ND4*, and *B2M* in a triplex qPCR reaction is an established method to determine mtDNA CN (Pyle *et al.*, 2010; Pyle *et al.*, 2015a; Pyle *et al.*, 2015b), therefore *B2M*, *MT-ND1*, and *MT-ND4* were measured in every UKBS sample (n=3091). Every qPCR run had a calibre control, a known deletion sample and a standard curve on the same plate with UKBS samples. Control DNA was the same as for PCR reactions (section 2.8.1.1). Calibre deletion DNA was extracted from trans-mitochondrial cybrids (provided by Professor CT Moraes, Miller School of Medicine, University of Miami, U.S.A.). The cybrids contained a single, large-scale mitochondrial deletion spanning the major arc from positions m.7982 to 15504 (Diaz *et al.*, 2002) at a heteroplasmy level of approximately 70%. Control and deletion DNA were diluted to 1/100 in water. To generate a standard curve, each standard template (*B2M*, *MT-ND1,* and *MT-ND4*) was freshly prepared and multiplexed in water to a single dilution of 1.00×10^9 copies/µl, and then serially diluted from 10^8 to 10^2 copies/µl. Standard curves were analysed simultaneously with calibre DNA and samples.

The triplex qPCR mastermix consisted of 0.2 µM of each TaqMan™ probe (Integrated DNA Technologies, Leuven, Belgium) (Table 2.4), 1x iTaq™ Universal Probes Supermix (BioRad, Hertfordshire, UK), and 0.3 µM of each forward and reverse primer (Integrated DNA Technologies, Leuven, Belgium) (Table 2.4). All reactions used Ambion® nuclease free water (Thermo Fisher Scientific, Loughborough, UK) and were set up on ice. qPCR products generated from the primers were *B2M*, *MT-ND1,* and *MT-ND4* (Table 2.4). 2 ng of DNA was loaded with 9 µl of mastermix on to each Hard-Shell®, thin wall 384 PCR plate (clear shell, white well), and was covered with a Microseal® 'B' seal plate seal (both BioRad, Hertfordshire, UK). All DNA samples, control and deletion calibres, and standards were measured in triplicate (Grady *et al.*, 2014b).

qPCR reactions were amplified on a CFX384 Touch™ Real-Time PCR Detection System (BioRad, Hertfordshire, UK), and the conditions were; initial denaturation 95°C for three minutes, followed by 40 cycles of denaturation at 95°C for 10 seconds, and annealing and extension at 62°C for one minute.

Gene	Amplicon Size (bp)	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')	Fluorophore	Quencher	Probe Sequence (5'-3')
B2M	231	CACTGAAAAAGATGAGTATGCC	AACATTCCCTGACAATCCC	FAM	BHQ_1	CCGTGTGAACCATGTGACTTTGTC
MT-ND1	111	ACGCCATAAAACTCTTCACCAAAG	GGGTTCATAGTAGAAGAGCGATGG	HEX	BHQ_1	ACCCGCCACATCTACCATCACCCTC
MT-ND4	107	ACCTTGGCTATCATCACCCGAT	AGTGCGATGAGTAGGGGAAGG	Cy5	BHQ_2	CAACCAGCCAGAACGCCTGAACGCA
MT-RNR2	126	CCAACGGAACAAGTTACCCTAG	CGAACCTTTAATAGCGGCTGC	HEX	BHQ_1	TACGACCTCGATGTTGGATCAGGACA

Table 2.4: A list of forward and reverse primer 5'-3' sequences and TaqMan™ probe sequences with their respective fluorophores for triplex quantitative PCR reactions. Genes amplified in all UKBS and 'Cohort Two' samples are nuclear gene beta-2-microglobulin (B2M) and mitochondrially encoded NADH-ubiquinone oxidoreductase core subunits 1 and 4 genes (MT-ND1 and MT-ND4 respectively). As detailed in section 5.4.6.2, mitochondrially encoded 16S ribosomal RNA (MT-RNR2) was quantified instead of MT-ND1. All primers and probes were synthesised by Integrated DNA Technologies in Leuven, Belgium.

Materials and Methods

59

2.8.1.5 Calculating Relative Mitochondrial DNA Copy number

After each qPCR run, initial analysis was performed using Bio-Rad CFX Manager 3.0 software. First, calibre DNA triplicates were checked for each amplicon, and replicates greater than 0.5 Ct difference were removed. For the first run, each fluorophore baseline threshold was set to an optimum, where standards were running parallel (Figure 2.2). Thresholds for *MT-ND1*-HEX and *MT-ND4*-Cy5 were set accordingly to *B2M* calibre Ct values. Average Ct values were noted for both control and deletion DNA calibre samples for each amplicon, and differences calculated. In every experiment following this, the reaction baseline threshold was then set to the control calibre DNA for each amplicon, and the difference between the average Ct for each amplicon for the control calibre was consistent (Table 2.5). Additionally, standard dilution replicate measurements were checked. The strongest and weakest standard dilution curves, and replicates greater than 0.5 Ct difference were removed to generate the most efficient standard curve for each fluorophore. A minimum of four standards dilutions were used for each amplicon, and samples had to sit within the selected standard curve range to have mtDNA CN calculated (Figure 2.3).

Standard curve efficiencies were only accepted above 90% for each fluorophore and negative control contamination above 35 Ct values. The Ct value for each standard dilution was not allowed to drift over one Ct from run to run as well, as this indicated standard templates were degrading (Table 2.5). Samples were analysed once thresholds were set and standard curves were running efficiently and parallel to each other. Sample replicates for each fluorophore were checked and replicates greater than 0.5 Ct difference were removed. Raw data was exported in to Microsoft Excel 2010, and mtDNA CN was calculated for each sample using Equation 2.2. Overall mtDNA CN calculated is an average of each individual replicate delta Ct value (Table 2.6), and is reported as copies per cell (copies/cell).

$$Relative\ Mitochondrial\ DNA\ Copy\ Number = 2(2\text{^-}\Delta\mathrm{Ct})$$

Equation 2.2: Calculation used to calculate mitochondrial DNA copy number (mtDNA CN) for each sample measured using qPCR. -∆Ct is the Ct difference between the mitochondrial reference gene (either MT-ND1, MT-ND4 or MT-RNR2) Ct and B2M Ct for each replicate. Overall mtDNA CN is an average of each 2(2^-∆Ct) replicate value. Relative mtDNA CN is recorded as copies/cell.

3028 UKBS samples (97.96%) successfully had relative mtDNA CN measured (n=1512 males, n=1516 females).

	Run 1	Run 2	Run 3
B2M - FAM			
LogSQ	Mean Ct	Mean Ct	Mean Ct
8.083			
7.083	14.31	14.71	14.69
6.083	17.68	18.36	18.83
5.083	21.11	22.01	22.17
4.083	24.84	25.20	25.47
3.083	28.36	28.44	28.94
2.083			
Negative	39.56	0	0
Threshold	110	110	110
Slope	3.533	3.430	3.512
RFU	4000	4000	4000
Efficiency	91.90%	95.70%	92.60%
Control (1/100)	28.68	28.63	28.89
Deletion (1/100)	26.44	26.75	26.63
MT-ND1 - HEX			
LogSQ	**Mean Ct**	**Mean Ct**	**Mean Ct**
8.114			
7.114			
6.114	17.14	17.42	17.92
5.114	20.55	20.75	21.00
4.114	23.98	24.09	24.26
3.114	27.33	27.72	27.71
2.114			
Negative	38.35	0	36.15
Threshold	116	102	110
Slope	3.4	3.448	3.261
RFU	3000	3000	3000
Efficiency	96.80%	95.00%	102.60%
Control (1/100)	22.81	22.76	23.02
Deletion (1/100)	19.07	18.82	18.78
MT-ND4 – Cy5			
LogSQ	**Mean Ct**	**Mean Ct**	**Mean Ct**
8.158			
7.158			
6.158	17.40	17.45	17.89
5.158	20.71	20.92	21.20
4.158	24.18	24.12	24.45
3.158	27.53	27.93	27.91
2.158			
Negative	37.63	0	0
Threshold	248	226	246
Slope	3.384	3.475	3.324
RFU	5000	5000	4000
Efficiency	97.50%	94.00%	99.90%
Control (1/100)	22.82	22.77	23.03
Deletion (1/100)	20.45	20.44	20.49
Control mtDNA - ND1	116.94	117.29	117.44
Deletion %	61.47%	67.42%	69.26%

Table 2.5: An overview of how standard optimum efficiencies were assigned and baseline thresholds were set to the control calibre DNA across three qPCR runs. The differences between control calibre Cts for each fluorophore is consistent between runs and is always set to B2M. This table also helped monitor standard template quality, deletion calibre quality and monitor overall run efficiency. RFU means relative fluorescent units.

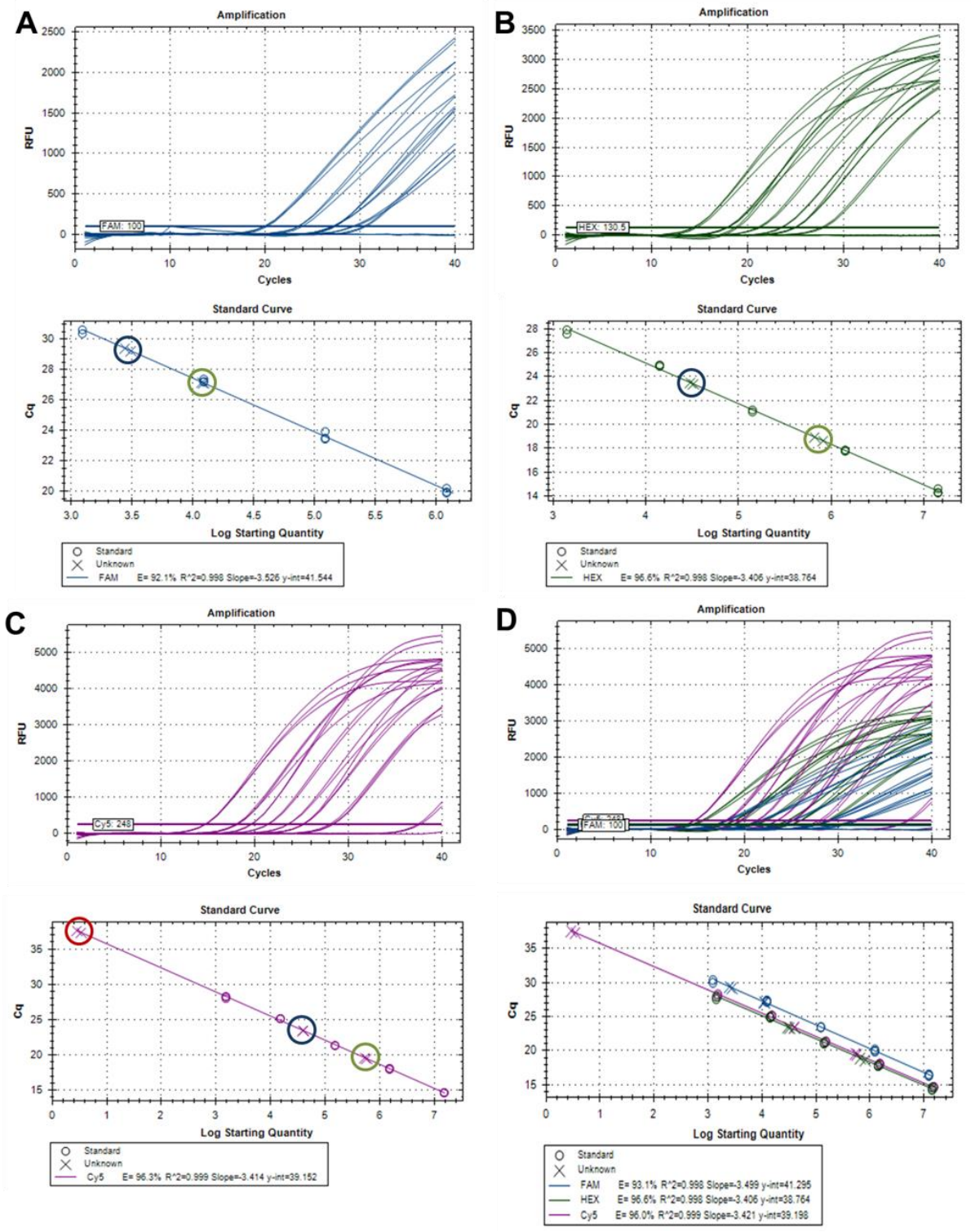


Figure 2.2: Images of amplification curves (top) and standard curves (bottom) for each respective fluorophore during quantitative real-time PCR analysis. [A] B2M detection using FAM. [B] MT-ND1 detection using HEX. [C] MT-ND4 detection using Cy5. [D] All three fluorophores overlaid. A-D show quantification of standard serial dilutions (small circles) and highlight where the control (large blue circle) and deletion (large green circle) calibre DNA amplify relative to the standard DNA. The large red circle in [C] depicts contamination >35 Cts. Analysis was performed on BioRad CFX 3.1 software.

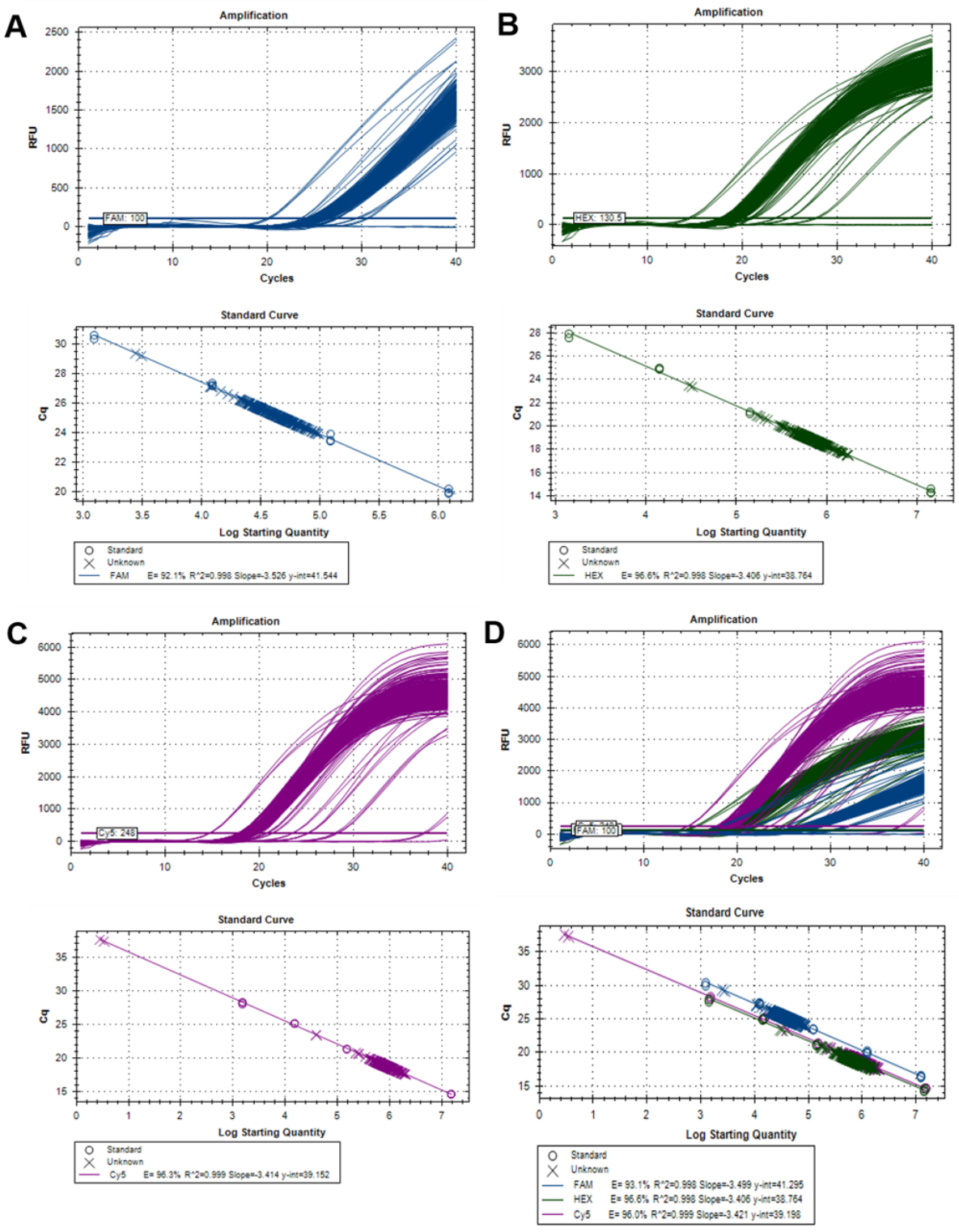


Figure 2.3: Images of amplification curves (top) and standard curves (bottom) for each respective fluorophore during quantitative real-time PCR analysis. [A] B2M detection using FAM. [B] MT-ND1 detection using HEX. [C] MT-ND4 detection using Cy5. [D] All three fluorophores overlaid. Images include amplification curves of UKBS samples (n=95) relative to the standard curves for each respective fluorophore. Analysis was performed on BioRad CFX 3.1 software.

		MT-ND1							B2M										
Run_ID	ID	Raw_Ct	Ct_Mean	Ct_St_Dev	SQ	Log_SQ	SQ_Mean	SQ_St_Dev	Raw_Ct	Ct_Mean	Ct_St_Dev	SQ	Log_SQ	SQ_Mean	SQ_St_Dev	ΔCT	2^-ΔCT	2(2^-ΔCT)	2(2^-ΔCT)
Run 1	Control (1/100)	23.607	23.701	0.088	51185.889	4.709	48209.382	2779.742	29.638	29.572	0.178	2528.725	3.403	2650.584	312.115	-6.031	65.380	130.760	
Run 1	Control (1/100)	23.714	23.701	0.088	47761.448	4.679	48209.382	2779.742	29.370	29.572	0.178	3005.246	3.478	2650.584	312.115	-5.656	50.423	100.846	
Run 1	Control (1/100)	23.782	23.701	0.088	45680.807	4.660	48209.382	2779.742	29.708	29.572	0.178	2417.782	3.383	2650.584	312.115	-5.925	60.774	121.547	117.718
Run 1	Deletion (1/100)	18.004	17.938	0.093	1952079.630	6.290	2039272.352	123309.130	26.568	26.458	0.156	18239.217	4.261	19626.382	1961.748	-8.565	378.628	757.255	
Run 1	Deletion (1/100)	17.872	17.938	0.093	2126465.074	6.328	2039272.352	123309.130	26.348	26.458	0.156	21013.548	4.322	19626.382	1961.748	-8.476	356.150	712.300	734.778
Run 1	Sample_1	19.860	20.066	0.178	584179.804	5.767	513390.737	61305.945	25.243	25.419	0.163	42803.689	4.631	38359.386	4073.396	-5.383	41.723	83.447	
Run 1	Sample_1	20.170	20.066	0.178	477680.234	5.679	513390.737	61305.945	25.450	25.419	0.163	37470.906	4.574	38359.386	4073.396	-5.280	38.848	77.696	
Run 1	Sample_1	20.168	20.066	0.178	478312.174	5.680	513390.737	61305.945	25.565	25.419	0.163	34803.562	4.542	38359.386	4073.396	-5.397	42.123	84.245	81.796
Run 1	Sample 2	20.205	20.278	0.083	467030.966	5.669	445897.204	23769.912	26.301	26.153	0.149	21661.230	4.336	23905.724	2288.431	-6.096	68.426	136.852	
Run 1	Sample 2	20.368	20.278	0.083	420163.525	5.623	445897.204	23769.912	26.154	26.153	0.149	23820.247	4.377	23905.724	2288.431	-5.786	55.182	110.364	
Run 1	Sample 2	20.260	20.278	0.083	450497.121	5.654	445897.204	23769.912	26.004	26.153	0.149	26235.696	4.419	23905.724	2288.431	-5.743	53.570	107.140	118.119

Table 2.6: An example of the table used to calculate relative mitochondrial DNA copy number for each UKBS sample using 2(2^-ΔCt) (Equation 2.2). Raw data for each sample from qPCR analysis was imported in to the table for B2M-FAM and MT-ND1-HEX and delta Ct calculations were applied respectively to each replicate. Relative mitochondrial DNA copy number was an average of each replicate's delta Ct calculation and is represented as copies/cell. Note – deletion calibre sample only has two replicates because one replicate was removed during QC. Analysis was performed in Microsoft Excel 2010.

2.8.1.6 Z-scoring

To normalise the distribution of mtDNA CN, and to accurately compare data between cohort populations in this study, Z-scoring was applied to each raw mtDNA CN value using Equation 2.3.

$$Z\ score = \frac{x - \mu}{\sigma}$$

Equation 2.3: Z-score calculation used to normalise mtDNA CN values. x is the raw mtDNA CN data value, μ is the population mean, and σ is the population standard deviation.

2.8.2 Cohort Two - In house 'Cohort Two'

Cohort two samples had mtDNA CN determined using exactly the same methods detailed in section 2.8.1, however 'Cohort Two' samples were measured in quadruplicate instead of triplicate measurements using qPCR (Table 2.8). Z-scoring mtDNA CN was also applied to mother and child cohorts separately.1854 (92.7%) samples had mtDNA CN determined.

2.8.3 Cohort Three – ALSPAC

5461 maternal ALSPAC samples had mtDNA CN assayed at the same time points blood cell proportions were estimated from methylation data (section 2.3.3) (Relton *et al.*, 2015). mtDNA CN was determined using an in-house qPCR assay which was optimised by Dr Philip A.I. Guthrie (PG) (who was formerly part of Dr Santi Rodriguez's group in the School of Social and Community Medicine at University of Bristol, UK). PG developed a singleplex assay which quantified the relative proportion of a region (bases 317 to 381) in the mitochondrial D-loop (GenBank: NC_012920.1) to nuclear reference gene *B2M* (chr 15, GenBank: NC_000015.10), using the established delta Ct calculation (section 2.8.1.5 - Equation 2.2).

For each qPCR reaction, a master mix was made using 5 μL of 2x SensiFAST SYBR® No-ROX Kit (Bioline, London, UK), 250 nM of each forward and reverse primer (Table 2.7), and 0.5 μL water (Malik *et al.*, 2011). 0.25 ng/μL of DNA was added to the mastermix to make a final volume of 10μL per well. Each mitochondrial and nuclear gene was measured in triplicate. Samples were assayed using a Roche LightCycler LC480 and 384-well plates under the following thermocycler conditions: one cycle for 5 minutes at 95^0C, followed by 45 cycles of: 5 seconds at 95^0C, 15 seconds at 55^0C, 15 seconds at 72^0C, and then 1 second at 78^0C. Standard curves

were generated from pooled results of quadruplicate DNA measurements, and were used to calculate reaction efficiencies, which were then used to adjust raw Ct data.

Three calibrator DNAs were run on every plate, and an average of each calibrator was calculated for each plate. This average was compared between plates, and was compared to the average of all of the calibrators across all plates, allowing the derivation of a per-plate calibration factor, by which each plate value was multiplied. This controlled for run-to run variability (Guyatt *et al.*, 2017). ALSPAC raw mtDNA CN values were then z-scored using Equation 2.3, and z-scored values were used in all further analysis. 5461 samples had mtDNA CN determined.

Gene	Amplicon Region (bp)	Forward Primer (5' – 3')	Reverse Primer (5' – 3')
B2M	4691 - 4717	GCTGGGTAGCTCTAAACAATGTATTCA	CCATGTACTAACAAATGTCTAAAATGGT
Mitochondrial D-loop	317 - 381	CTTCTGGCCACAGCACTTAAAC	GCTGGTGTTAGGGTTCTTTGTTTT

Table 2.7: Details of forward and reverse quantitative PCR primers used to quantify nuclear gene B2M and mitochondrial D-loop region in ALSPAC samples. Assay optimised by Dr Philip A.I. Guthrie (PG) (who was formerly part of Dr Santi Rodriguez's group in the School of Social and Community Medicine at University of Bristol, UK).

Cohort #	Cohort Name	Housekeeping nDNA Gene	nDNA Amplicon Region	Mitochondrial mtDNA Genes	mtDNA Amplicon Regions	Triplex/ Singleplex	# Replicates Measured	qPCR Probes	Results Chapter	Notes
1	UKBS	*B2M*	chr15; 9145-9375	*MT-ND1, MT-ND4*	3458-3568, 11144-11250	Triplex	3	TaqMan™	All	
		B2M	chr15; 9145-9375	*MT-RNR2, MT-ND4*	2911-3036, 11144-11250	Triplex	4	TaqMan™	3	Used as an 'alternative' qPCR assay to validate results generated using *MT-ND1* assay
2	In house 'Cohort Two'	*B2M*	chr15; 9145-9375	*MT-ND1, MT-ND4*	3458-3568, 11144-11250	Triplex	4	TaqMan™	1 and 3	
		B2M	chr15; 9145-9375	*MT-RNR2, MT-ND4*	2911-3036, 11144-11250	Triplex	4	TaqMan™	3	Used as an 'alternative' qPCR assay to validate results generated using *MT-ND1* assay
3	ALSPAC	*B2M*	chr15; 4691-4717	D loop	317 - 381	Singleplex	3	Sybr Green®	2 and 3	

Table 2.8: Summary of quantitative real-time PCR methods used to quantify mitochondrial DNA copy numbers in samples from each cohort throughout this study, and details of when each qPCR assay was used to generate results for each respective results chapter. Cohorts 1 and 2 were measured at Newcastle University, UK. Cohort 3 was measured at University of Bristol, UK and data was used in collaboration with Dr Santi Rodriguez.

2.9 Investigating Quantitative PCR Reproducibility

qPCR is a sensitive method, and a number of factors can contribute to variation recorded between replicates, including the quality of reagents or DNA, evaporation, and pipetting error. This variation contributes to differences in overall final mtDNA CN calculations. Therefore, before embarking on measuring relative mtDNA CN in a large cohort of samples (n=3091), the number of replicate qPCR measurements was assessed and established in a smaller pilot cohort (n=8) to ensure precise mtDNA CN values were recorded throughout the project. In addition, mtDNA CN consistency, calculated using either *MT-ND1* or *MT-ND4* was assessed between replicate measurements.

2.9.1 Identifying the Most Suitable Number of qPCR Replicates

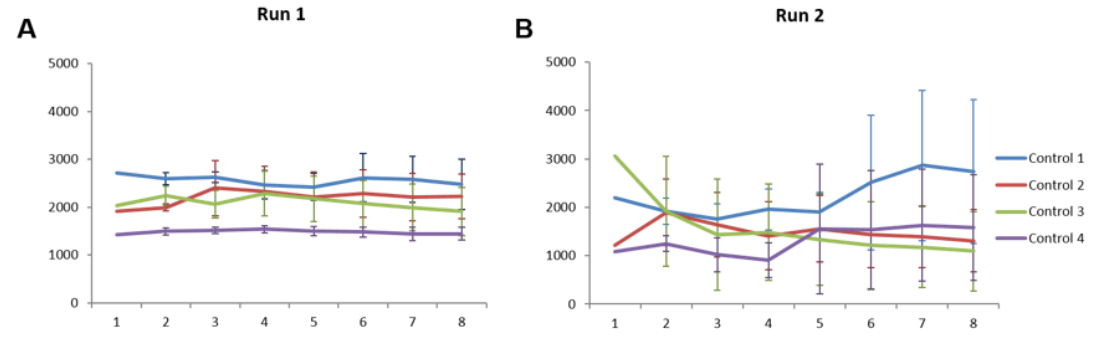
Eight independent samples of control DNA (detailed in section 2.8.1.1) were diluted to 1/10 (approximately 70 ng/µl) and were run in replicates of eight against triplicates of standard DNA and calibre deletion DNA (detailed in section 2.8.1.4) on a Hard-Shell®, 96 well PCR plate (white well, green shell) (BioRad, Hertfordshire, UK). The entire experiment was repeated twice. Relative mtDNA CN was then calculated for each individual replicate for each control sample using Equation 2.2 (Pyle *et al.*, 2007). mtDNA CN was calculated using either *MT-ND1* or *MT-ND4* relative to *B2M* to compare mtDNA CN differences dependent on the mitochondrial reference gene.

To investigate the variability in mtDNA CN when including a different number of averaged replicates in the delta Ct calculation, the cumulative average and standard deviation for each number of replicates was calculated and compared. For example; for three replicates, the average of the first three mtDNA CN measurements was recorded and then the standard deviation calculated to identify the amount of variation between those three individual mtDNA CN measurements. This was repeated where one to eight replicates were compared to calculate mtDNA CN in the delta Ct calculations (Figure 2.4). Tables a-d in Figure 2.4 summarise the average mtDNA CN and standard deviation across eight control samples calculated using the respective number of replicates.

Figure 2.4 shows a general pattern where standard deviations start to decrease as more replicates are included to calculate mtDNA CN, then the standard deviations plateau between three to five replicates, and then they start to increase again as more replicates are included.

As advised by Robert Kitchen (Kitchen *et al.*, 2010) - no fewer than three technical replicates should be used when performing qPCR measurements on biological samples. Additionally, he states that increasing the number of replicates only improves measurement precision, and you need to increase sample numbers to record true biological differences (Kitchen *et al.*, 2010). Following this advice, and in conjunction with the results in Figure 2.4, we measured samples in triplicate or quadruplicate (Table 2.8).

MT-ND1: B2M



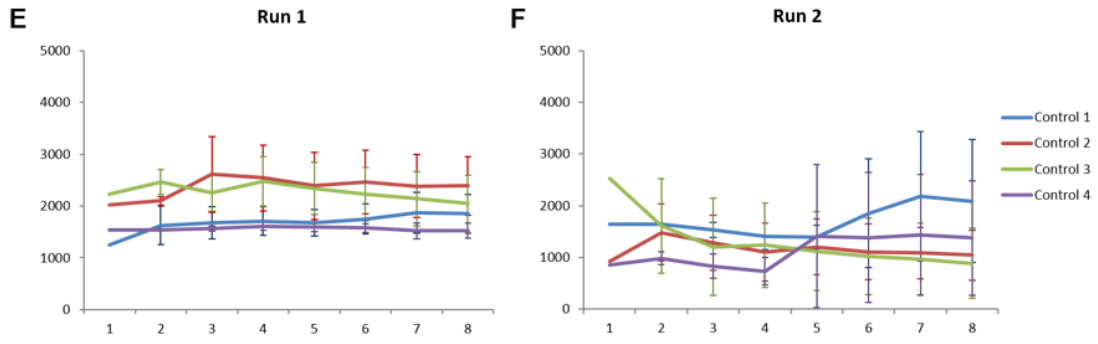
a

	1	2	3	4	5	6	7	8
Mean	1917.78	1976.07	2024.08	2009.19	1986.01	1997.74	1952.99	1945.90
St Dev	411.74	405.86	454.49	422.88	395.77	420.21	416.46	382.42

b

	1	2	3	4	5	6	7	8
Mean	1966.69	1773.15	1726.35	1730.13	1730.07	1768.67	1805.78	1905.13
St Dev	892.06	306.68	479.78	510.71	247.46	401.37	512.03	582.53

MT-ND4: B2M



c

	1	2	3	4	5	6	7	8
Mean	1837.38	1942.84	1990.91	2001.16	1975.80	1969.01	1932.68	1931.74
St Dev	370.49	407.68	456.24	440.59	399.29	372.83	354.24	335.43

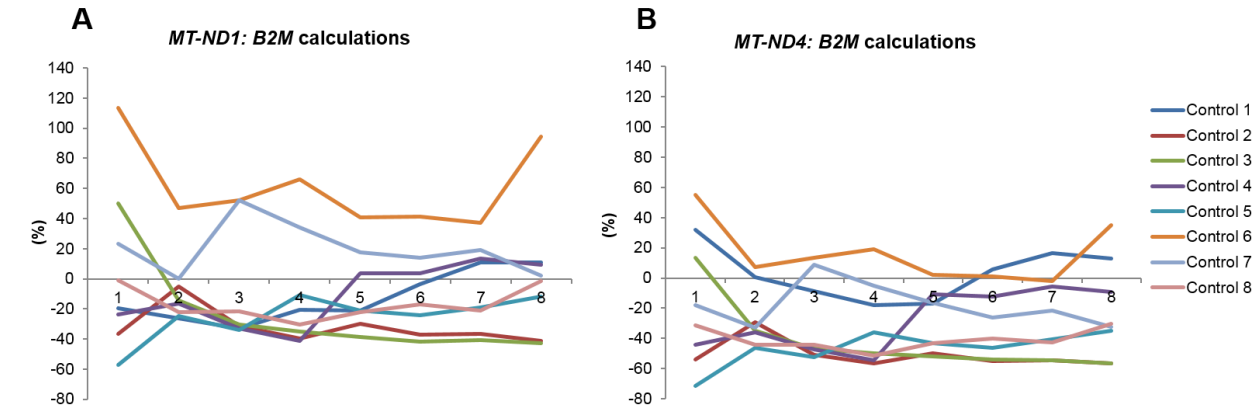
d

	1	2	3	4	5	6	7	8
Mean	1496.96	1380.19	1349.05	1301.81	1339.06	1339.06	1373.83	1435.78
St Dev	720.59	273.60	357.14	356.92	148.58	253.27	366.90	394.58

Figure 2.4: Graphs showing mitochondrial DNA copy number (mtDNA CN) changes (copies/cell) (y-axis) depending on the number of cumulative averged replicates included in the delta Ct calculation (x-axis), and the consistency between two separate qPCR runs. [A] – [D] use MT-ND1: B2M to calculate mtDNA CN. [E] - [H] use MT-ND4: B2M to calculate mtDNA CN. Tables a-d summarise means and standard deviations of mtDNA CN calculated for each number of replicates across eight control samples measured.

2.10 Determining which Mitochondrial Reference Gene to Use to Calculate Mitochondrial DNA Copy Number

To determine which mitochondrial reference gene would be more appropriate to calculate relative mtDNA CN, and elaborating on the work in section 2.9, relative mtDNA CN was calculated for each number of replicates from the eight controls using either *MT-ND1* or *MT-ND4* in Equation 2.2. After repeating the assay, the difference in mtDNA CN using either gene was calculated and compared (Figure 2.5).



	1	2	3	4	5	6	7	8
Mean	6.16	-7.68	-9.91	-9.72	-8.87	-8.08	-4.68	2.25
St Dev	55.15	23.98	38.42	39.14	27.06	27.68	28.54	42.65

	1	2	3	4	5	6	7	8
Mean	-14.82	-26.88	-28.46	-31.62	-28.78	-28.33	-25.63	-21.56
St Dev	44.32	19.94	28.06	27.60	20.55	24.32	26.60	32.35

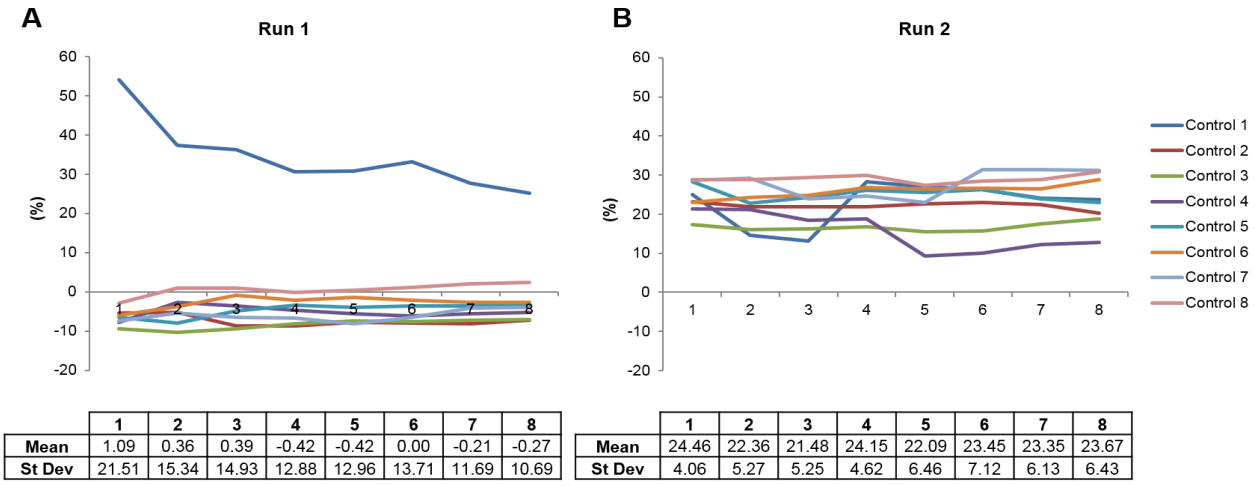
Figure 2.5: Graphs reporting the percentage differences in relative mitochondrial DNA copy number (mtDNA CN) (y axis) between two experiments when using the cumulative average of different number of replicates (x axis). [A] Uses MT-ND1: B2M to calculate mtDNA CN. [B] Uses MT-ND4: B2M to calculate mtDNA CN. Tables summarise the average differences in mtDNA CN (y axis) between runs per number of replicates included across eight control samples, and their standard deviations. This was repeated across two separate qPCR runs.

The average difference between qPCR runs, when using *MT-ND1: B2M* calculations was -5.07% (±35.33 s.d) (Figure 2.5A), whereas the average difference when using *MT-ND4: B2M* calculations was -25.76% (±27.97 s.d) (Figure 2.5B). This suggests *MT-ND4* is more variable between runs. Although these values are high, similar differences between the *MT-ND1* and *MT-ND4* calculations have been reported in another study, although they only reported a 2.5% difference (He *et al.*, 2002).

MT-ND4 is extensively researched and is known to be susceptible to mutations and deletions, therefore I chose to use *MT-ND1* to calculate mtDNA CN, to generate more consistent results and keep in line with previously optimised qPCR assays (He *et al.*, 2002; Payne *et al.*, 2011; Yu-Wai-Man *et al.*, 2011; Phillips *et al.*, 2014; Rygiel *et al.*, 2015).

2.10.1 Run to Run Variability

As reported above (section 2.9), run to run variation also affects mtDNA CN reproducibility. As shown in Figure 2.4, mtDNA CN values were not replicated, irrespective of the number of replicates used. This may be due to technical error between runs which causes inconsistent quantification of each gene being measured, leading to variation between mtDNA CN values calculated. Figure 2.6 confirms this and shows the differences in mtDNA CN between runs when comparing *MT-ND1* and *MT-ND4* genes.



	1	2	3	4	5	6	7	8
Mean	1.09	0.36	0.39	-0.42	-0.42	0.00	-0.21	-0.27
St Dev	21.51	15.34	14.93	12.88	12.96	13.71	11.69	10.69

	1	2	3	4	5	6	7	8
Mean	24.46	22.36	21.48	24.15	22.09	23.45	23.35	23.67
St Dev	4.06	5.27	5.25	4.62	6.46	7.12	6.13	6.43

Figure 2.6: Graphs reporting the percentage differences in relative mitochondrial DNA copy number (mtDNA CN) values when mtDNA CN is calculated using MT-ND1: B2M and MT-ND4: B2M (y axis) calculations. mtDNA CN was calculated using a cumulative average of a specific number of replicates (x axis).Tables summarise the average differences in mtDNA CN (y axis) per number of replicates included across eight control samples, and their standard deviations. This was repeated across two separate qPCR runs.

The average difference between *MT-ND1: B2M* calculated or *MT-ND4: B2M* calculated mtDNA CN values, across all replicates; in run 1 was 0.06% (±14.22 s.d). However, excluding the outlier sample (Figure 2.6A, blue line), the average difference was -4.84% (±3.25 s.d). For run 2 (Figure 2.6B), the average difference across all replicates was 23.12% (±5.67 s.d). This demonstrated that run-to-run variability could range between 5% and 23% depending on the number of replicates measured, and the reference gene being used to calculate mtDNA CN.

To investigate this further, a random selection of UKBS samples had their mtDNA CN measured across several different qPCR runs. To identify the real technical error contributing to noise, samples were measured in triplicate and relative mtDNA CN was calculated using delta Ct of *MT-ND1: B2M* because there was less variability in mtDNA CN between runs (Figure 2.5A).

35 UKBS samples were measured across two separate experiments, and 59 UKBS samples were measured across three separate experiments. The mtDNA CN generated across the two-run tests varied by an average of 16%, with differences ranging from 1-34%. The mtDNA CN generated across the three-run tests varied by an average of 18%, with differences ranging from 1-45%. These data, combined with that reported above, suggests there is an approximately 20% degree of run-to-run variation in mtDNA CN values from accumulated technical error. As a consequence of this, any results will need to factor a 20% difference in to be significant; otherwise the result may be due to technical variability.

2.11 Power Calculations

Statistical power calculations were performed using a G×G or G×E interaction, Windows-based software program called QUANTO (Morrison and Gauderman, 2002).

Based on previous publications, mean mtDNA CN calculated from genomic DNA extracted from blood ranges from 50 to 400 copies/cell, with standard deviations ranging from 5 to 37% (Gahan *et al.*, 2001; Xing *et al.*, 2008; Phillips *et al.*, 2014; Ding *et al.*, 2015; Pyle *et al.*, 2015a; Guyatt *et al.*, 2017). mtDNA CN heritability is estimated at 65% (Xing *et al.*, 2008). Being conservative, power calculations were generated on the assumption that mean mtDNA CN would be 200 copies/cell, with a standard deviation of 15% (equating to ±30 copies/cell).

Calculations used a continuous trait study design, based on an additive, gene-only effects hypothesis in independent individuals. Allele frequencies between 0.01 and 0.51, with an effect size (β_G) of 0.8 to 6.0 were simulated to a type one error rate of 5.00E-08 (genome-wide significance) to detect power. Based on the UKBS cohort size (n=3091), there is 83.52% power to detect a nuclear variant with a β_G of 6.0, with a minor allele frequency (MAF) of 21%. As mtDNA CN variability is common, I expect to detect a variant with a higher allele frequency. This study has 80.45% power to detect a nuclear variant with a β_G of 4.8, with a MAF of 51% (Figure 2.7).

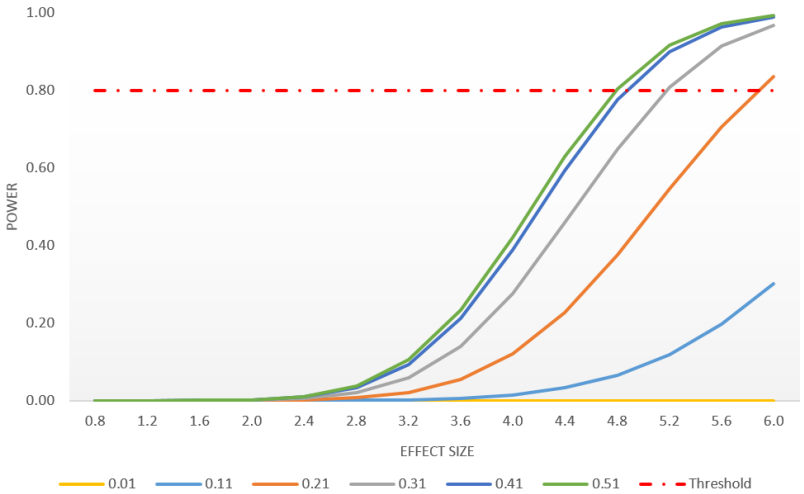


Figure 2.7: Simulated power calculations to determine a genome-wide significant (p=5.00E-08) nuclear variant associated with mitochondrial DNA copy number (μ=200 copies/cell ±30 copies) in UKBS samples (n=3091). Individual lines represent variant minor allele frequencies. Red dotted line indicates 80% power threshold.

Chapter 3 Mitochondrial DNA Copy Number in European Populations

3.1 Introduction

Mitochondrial DNA (mtDNA) was originally discovered in chick embryos using electron cytochemical techniques (Nass and Nass, 1962; Nass and Nass, 1963b), where two to six copies of mtDNA were yielded from each mitochondrion (Nass, 1969). The number of mtDNA in every cell, across a number of tissues and species, was originally thought to be the same (Nass *et al.*, 1965; Schneider and Kuff, 1965; Parsons and Simpson, 1967), however, in 1990, Karen Veltri and colleagues challenged this, reporting distinctly different numbers of mtDNA across a number of organs from mice (Veltri *et al.*, 1990). The advance of research since then has identified that mitochondria contain tens to thousands of mtDNA molecules per mitochondrion (Garrido et al., 2003; Bogenhagen et al., 2008). However, beyond energetic demand (Dickinson *et al.*, 2013; Fukuoh *et al.*, 2014), we know little of why mtDNA CN varies between individuals at the organism, tissue or cellular level, or what other factors modulate mtDNA levels.

mtDNA is packaged into nucleoids which are approximately 100 nm in size (Ylikallio *et al.*, 2010), and can contain between 1-10 copies of mtDNA, but this depends on cellular metabolic demand, the stage of cell cycle, and redox state within the cell (Garrido *et al.*, 2003; Bogenhagen *et al.*, 2008; Malik *et al.*, 2011; Malik and Czajka, 2013). For example, high energy-demanding cells such as neurons can contain around 2000 mitochondria (Uranova *et al.*, 2001), containing tens of thousands of mtDNA molecules (Dickinson *et al.*, 2013). However, different regions of the brain contain varying mtDNA CN (Fuke *et al.*, 2011; Rice *et al.*, 2014). Less energy-demanding cells, such as white blood cells, contain less than one hundred mitochondria, which can contain up to approximately 50 mtDNA molecules per cell (Selak *et al.*, 2011). mtDNA CN is considered an important indicator of mitochondrial activity because mitochondrial gene transcripts are proportional to mtDNA CN (Williams, 1986).

In NCBI PubMed, a search of [mitochondrial copy number] yields >2000 publications. Within the literature it is becoming more apparent the importance of understanding mitochondrial maintenance and mtDNA CN regulation, especially as mtDNA CN is thought to be a biomarker of disease susceptibility (Malik *et al.*, 2011; Giordano *et al.*, 2014; Lee *et al.*, 2014b; Pyle *et al.*, 2015b; Wei *et al.*, 2017), progression, and severity (Malik and Czajka, 2013). In mice, increased mtDNA CN has been seen to be detrimental, as it causes increased nucleoid enlargement, leading to defective

transcription, and results in the accumulation of mtDNA deletions, and OXPHOS defects (Ylikallio *et al.*, 2010). However, this may have been a result of the mouse model's phenotype overexpressing *Twinkle* or *TFAM*, continually stressing mitochondria to replicate. Needless to say, there has been a growing importance to understand the regulation of mtDNA CN; however no study has investigated mtDNA CN differences in large populations before. This is the first study to address this.

3.1.1 Principles of Quantitative-PCR

Since the introduction of quantitative PCR (qPCR) in 1996 (Heid *et al.*, 1996), qPCR has become the established method to determine mtDNA CN (Wai *et al.*, 2008; Malik *et al.*, 2011; Grady *et al.*, 2014a; Pyle *et al.*, 2015a; Rygiel *et al.*, 2015; Pyle *et al.*, 2016). Previous methods used radiolabelled hybridisation to quantify mtDNA (Veltri *et al.*, 1990; Tang *et al.*, 2000; Rodriguez-Enriquez *et al.*, 2009), however qPCR is quicker, more widely available, and requires less sample, and is therefore generally preferred in practice (Malik and Czajka, 2013).

The principle of qPCR is similar to that of PCR, whereby a sample of DNA acts as a template and undergoes a series of cyclic temperature conditions to; denature DNA, allow specific primers to bind, amplify the specific region of interest, then dissociate and allow DNA to anneal and extend. qPCR is different to PCR because quantification of the amplified product being measured is in real-time, and is quantified after each cycle across a number of PCR cycles, usually between 25 to 40 (Rebouças *et al.*, 2013; Ma and Chung, 2014). Relative mtDNA CN is calculated by measuring the proportion of a nuclear housekeeping gene (to adjust for cell number) to a mitochondrial reference gene, also known as the Mt/N ratio (Malik *et al.*, 2011).

Across the literature, a number of different nuclear housekeeping genes are used. Globulins (especially β-globin), glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), ubiquitin C (*Ubc*), β-actin (*ACBT*), β-tubulin, and 18S ribosomal protein (18S) are a select few (Miller *et al.*, 2003; Shen *et al.*, 2010; Albershardt *et al.*, 2012; Kelly *et al.*, 2012; Ridge *et al.*, 2014). Selecting a suitable housekeeping gene for qPCR normalisation has been highly debated, and is both tissue and species specific (Kelly *et al.*, 2012). For example, in one study, Kelly and colleagues reported that β-actin and *GAPDH* varied, whilst 18S was consistent across several different mouse cell and tissue types (Kelly *et al.*, 2012). However, Albershardt and colleagues argued *Ubc* was the most stable housekeeping gene in mouse lymphocytes

(Albershardt *et al.*, 2012). *B2M* has been identified as the most suitable housekeeping gene to determine mtDNA CN in human tissue because it is present in a single copy, and has little variability (Malik *et al.*, 2011; Venegas and Halberg, 2012; Phillips *et al.*, 2014; Rygiel *et al.*, 2015), therefore I used *B2M* as the housekeeping gene in this work.

MT-ND1 is established as the mtDNA CN reference gene used to measure mtDNA CN as it is less prone to mtDNA deletion (Krishnan *et al.*, 2007; Shen *et al.*, 2010; Pyle *et al.*, 2015b). However, other studies have measured regions in the mitochondrial D-loop (Phillips *et al.*, 2014), cytochrome B (Malik *et al.*, 2009), tRNA-Leu, and COX-IV (Venegas and Halberg, 2012; Rooney *et al.*, 2015). The mitochondrial D-loop is generally avoided because it is a triple stranded region, thought to regulate mtDNA replication, and contains hypervariable regions which can interrupt PCR amplification machinery (Anderson *et al.*, 1981; Takamatsu *et al.*, 2002; J Nicholls and Minczuk, 2014; Jemt *et al.*, 2015). Careful consideration needs to be made when designing mitochondrial amplicons, as nuclear-mitochondrial DNA segments (NUMTs) in nDNA may cause co-amplification of nDNA, which can affect the mtDNA CN outcome (Hazkani-Covo *et al.*, 2010; Calabrese *et al.*, 2012).

The efficiency and consistency of qPCR results are dependent on a number of factors, including: reagents, such as primers and probes, DNA quality and source, size of the amplicon being amplified, and the number of replicates used. For example, single stranded breaks in mtDNA interrupt binding of DNA polymerases (Ponti *et al.*, 1991; Gonzalez-Hunt *et al.*, 2016), and freeze-thawing DNA samples can reduce the quality of DNA (Ross *et al.*, 1990), emphasising the importance of DNA quality for qPCR measurements. Triplicate measurements are also the most common number of replicates used in qPCR, and as my work in section 2.9.1 showed, mtDNA CN differs across varying replicate sizes. However, some studies only use duplicate measurements (Shen *et al.*, 2010). Gervais and colleagues also showed that qPCR protocol temperature, DNA amplicon size, and amplicon region all affect primer efficiencies (Gervais *et al.*, 2010).

In qPCR, SYBR® green has traditionally been used to measure mtDNA CN because it provides a quick and convenient master mix of all the components (except primers) needed for qPCR, along with a fluorophore that binds to double-stranded DNA, and is easily detectable (Applied Biosystems, 2002; Fuke *et al.*, 2011; Chan *et al.*, 2013;

Rooney *et al.*, 2015; Gonzalez-Hunt *et al.*, 2016). However, the introduction of specific 5'-3' exonuclease technology TaqMan™ probes in 1991, meant several primer-specific and probe-specific reagents could be multiplexed together to detect multiple genes at once. TaqMan™ probes are also more specific than SYBR® green as they target a specific area of the amplicon whereas SYBR® green targets any sequence of double stranded DNA. As a result, TaqMan™ probes revolutionised qPCR because it meant both mtDNA CN and mitochondrial deletions could be detected in one measurement, and with a higher degree of accuracy (Holland *et al.*, 1991; Nicklas *et al.*, 2004; Grady *et al.*, 2014a; Phillips *et al.*, 2014). What's more is recently droplet-digital PCR (ddPCR) is becoming more commonly used, because it can detect low levels of mtDNA, which SYBR® green and TaqMan™ cannot (Henrich *et al.*, 2012; Taylor *et al.*, 2014; Belmonte *et al.*, 2016; Ye *et al.*, 2017). On the contrary, some studies have also used exome or next generation sequencing to calculate mtDNA CN, by calculating the ratio of aligned mitochondrial reads to aligned nuclear reads. However, this is under the assumption that both nuclear and mitochondrial regions of the genome are of equal ploidy, and are sequenced to comparable depths, and is not an actual measure of mtDNA CN (Guo *et al.*, 2013; D'Erchia *et al.*, 2014; Cai *et al.*, 2015; Ding *et al.*, 2015; Reznik *et al.*, 2016).

In addition to qPCR technical limitations, the source of genomic DNA also affects mtDNA CN outcome. For example, saliva is a readily accessible, and much cheaper, source of genomic DNA. However, it is less commonly used in research studies because little is known about the source of the DNA (whether it be host, or contaminated genomes from food or bacterial sources (Takeshita *et al.*, 2016)), and larger volumes of sample are needed to yield sufficient DNA (Ridge *et al.*, 2014). 74% of DNA in saliva is believed to come from white blood cells, with leukocyte levels ranging from 2 to 1.36E+05 cells/ml, and up to 1.10E+06 cells/ml in patients with oral cavity inflammation (Thiede *et al.*, 2000). However, white blood cells, extracted from whole blood, are still the preferred source of genomic DNA for studies as they are known to contain a large amount of DNA.

3.1.2 Other Confounding Factors Affecting Mitochondrial DNA Copy Number

3.1.2.1 DNA Quality Effect on Mitochondrial DNA Copy Number

The median and mean are central tendency statistical values which provide information about population data. Statisticians debate whether it is more accurate to

use median values instead of the mean, because the median is a resistant measure and is less likely to be affected by outliers. However, the median does not give a clear indication of the spread of data across the population like the mean does (LeBlanc, 2004). The established method for calculating mtDNA CN uses the mean of triplicate or quadruplicate values generated from qPCR measurements (Amaral *et al.*, 2007; Phillips *et al.*, 2014; Rooney *et al.*, 2015; Pyle *et al.*, 2016), however no study has compared median or mean-derived mtDNA CN values before.

Aim 1: Identify if there are statistical differences between median and mean-derived mtDNA CN values using a large, control, European cohort.

3.1.2.2 Gender Effect on Mitochondrial DNA Copy Number

The maternal inheritance of mtDNA has biologically burdened the health of males, and is considered a sex-specific selective sieve (Camus *et al.*, 2012). This is because males inherit mtDNA which may contain deleterious effects that will negatively impact male metabolic systems, that female systems can overcome (Camus *et al.*, 2012; Frank, 2012). As a result, mitochondrial biogenesis has been reported to be different between sexes, including in rodents (Khalifa *et al.*, 2017; Mauvais-Jarvis *et al.*, 2017).

Males have more pronounced metabolic phenotypes than females (Mauvais-Jarvis *et al.*, 2017) PD, and amyotrophic lateral sclerosis (ALS) are male predominant, as is the primary mitochondrial disease LHON (Man *et al.*, 2003; Van Den Eeden *et al.*, 2003; Kim *et al.*, 2012). Moreover, mtDNA CN has been reported to be significantly lower in males with diseases such as PD (Pyle *et al.*, 2015a), yet has been reported to be higher in male cancer patients compared to females (Yin *et al.*, 2004; Thyagarajan *et al.*, 2012; Shen *et al.*, 2016). Several studies have reported mtDNA CN differences between genders, however no study has directly looked at mtDNA CN differences between genders in large, healthy, European populations.

Aim 2: Investigate whether mtDNA CN differs between genders in a large, healthy, European population.

3.1.2.3 Age Effect on Mitochondrial DNA Copy Number

mtDNA is more likely to accumulate DNA lesions with age because, unlike nDNA, mtDNA lacks nucleotide excision repair mechanisms, and is exposed to higher levels of ROS (Harman, 1956; Ledoux *et al.*, 1992). Nearly 5% of oxygen consumed is

converted into ROS, meaning all tissues are susceptible to age-related damage (Malik and Czajka, 2013). The accumulation of mtDNA defects is reported to decrease mtDNA replication and transcription, resulting in mtDNA CN changes (Niranjan *et al.*, 1982; Trifunovic *et al.*, 2004; Meyer *et al.*, 2013). This has led to age-related mtDNA CN changes being reported in a number of different tissues across species (Tezze *et al.*; Barrientos *et al.*, 1997; Barazzoni *et al.*, 2000; Miller *et al.*, 2003; Cree *et al.*, 2008a; Cheng and Ivessa, 2010; Payne *et al.*, 2011; Mengel-From *et al.*, 2014; Laubenthal *et al.*, 2016). T2DM, cancer, and PD are all age-related diseases, and are tissue specific (Singh, 2004; Bender *et al.*, 2006; Shen *et al.*, 2010; Pringsheim *et al.*, 2014), and mtDNA CN changes have also been observed in these diseases (Malik *et al.*, 2009; Shen *et al.*, 2010; Pyle *et al.*, 2015a; Pyle *et al.*, 2015b).

Additionally, a study which introduced deleterious mtDNA in *Drosophila melanogaster* showed accelerated aging in males compared to females (Camus *et al.*, 2012), and primary mitochondrial-related diseases such as LHON are more prevalent in males, where age of onset is in late teens (Wallace *et al.*, 1988; Man *et al.*, 2003; Hudson *et al.*, 2007b). What's more is Giordano's team reported that individual's carrying LHON mutations who do not develop the disease, known as incomplete penetrance, also have a higher mtDNA CN (Giordano et al., 2014). These data suggest that there may be a gender-specific effect causing mtDNA CN to change with age, and there may be compensatory mechanisms to protect against mitochondrial genetic insults relative to gender.

Aim 3: Identify the impact ageing has on mtDNA CN changes in a large, healthy, European population, and investigate whether ageing effects are gender-specific.

3.1.2.4 Blood Composition Effect on Mitochondrial DNA Copy Number

Peripheral blood cell count differs, depending on age, gender, and ethnicity. For example, Bellamy and colleagues report significant changes in white blood cells (WBC) subcomponent composition in European children between the ages of two to 13 months of age (Bellamy *et al.*, 2000), and Bain and colleagues also reported significantly lower WBC subcomponent cells in peripheral blood of African and Afrocarribean people compared to Caucasian people (Bain *et al.*, 1984; Bain, 1996).

Changes in WBC in peripheral blood are used as a prognosis index to determine host immunity, and the ratio of lymphocytes to neutrophils is a common marker (Walsh *et al.*, 2005; Gwak *et al.*, 2007). Neutrophils and macrophages are thought to

play a pivotal role in cancer tumour progression, whereby elevated neutrophil and monocyte counts in patients with metastatic renal cell carcinoma are considered a poor survival prognosis factor, as well as elevated LDH levels (Schmidt *et al.*, 2005). This may be because activation of monocytes results in excessive production of cytokines and other inflammatory mediators as the immune system is primed, which can act as a 'first hit' in healthy individuals. However, this is disadvantageous as any further injury like sepsis results in a hugely amplified inflammatory response, which leads to cell lysis, and the production of cytokine poisoning (O'Sullivan and O'Connor, 1997).

Sepsis is a rare, but serious condition which can lead to organ failure and death. Sepsis is very much dependent on immune cellular response, and mtDNA CN changes dependent on blood cell proportions (Pyle *et al.*, 2010). mtDNA is believed to be able to influence the immune system through inflammasomes, because NLRP3 inflammasome activity is regulated by ROS, which damaged mitochondria are known to generate (Gurung *et al.*; Zhou *et al.*, 2011; Cristina Kenney *et al.*, 2014; Yu and Lee, 2016; Harrington *et al.*, 2017). As a result, mtDNA CN has been reported to vary in cellular components of blood in patients with sepsis (Lorente *et al.*, 2013), and has been reported to change in peripheral blood from patients with metabolic diseases such as T2DM, cancer, and obesity (Malik and Czajka, 2013; Lee *et al.*, 2014b; Reznik *et al.*, 2016; Cho *et al.*, 2017). Additionally, a decrease in mtDNA CN in peripheral blood of 1700 control Chinese individuals was also associated with a 50% higher cardiovascular risk (Huang et al., 2016).

As the literature suggests, mtDNA CN in peripheral blood may act as both a biomarker of disease susceptibility, as well as disease progression, and may also be a contributing factor to cellular composition changes.

Aim 4: Identify if mtDNA CN changes in individual blood cell subcomponents, and investigate how mtDNA CN changes dependent on blood cell composition in a large, healthy, European population.

3.1.2.5 Inheritance of Mitochondrial DNA Copy Number

mtDNA is strictly inherited down the maternal line in humans (Birky, 1978; Giles *et al.*, 1980), because mtDNA in paternal sperm is degraded by the oocytes autophagy mechanisms soon after fertilisation (Al Rawi *et al.*, 2011; Sato and Sato, 2011; Politi *et al.*, 2014). This results in the oocyte containing thousands (10^5) of copies of

maternal mtDNA (Michaels *et al.*, 1982), of which all copies contain the same wildtype genome (homoplasmic), or are a mixture of wildtype and mutated mtDNA (heteroplasmy). Tissues are normally able to withstand up to a heteroplasmy threshold of 60-80% before phenotypes show, however this is dependent on the mtDNA mutation and tissue (Sciacco *et al.*, 1994; Wallace and Chalkia, 2013; Stewart and Chinnery, 2015). Women produce oocytes which contain widely different levels of heteroplasmy (Wallace and Chalkia, 2013; Hyslop *et al.*, 2016), and new therapies, such as pronuclear transfer, have recently been introduced to prevent the transmission of severely heteroplasmic pathogenic variants to offspring (Craven *et al.*, 2017).

Natural selection for mtDNA heteroplasmy genotypes is known as the mitochondrial bottleneck. The bottleneck mechanism is still undetermined; however several theories have been hypothesised. Cao and colleagues believe heteroplasmy is transmitted in mtDNA clusters of identical molecules, and non-mutant mtDNA is selected for as a result of preferential replication (Cao *et al.*, 2007; Cao *et al.*, 2009). Cree and colleagues however suggest genetic segregation is a result of significantly decreased mtDNA CN during early germ line development (Cree *et al.*, 2008b). Wai and colleagues challenge this theory by reporting continued high mtDNA CN in oocytes, and argue that the bottleneck effect is a result of replicating subpopulations of mtDNA during postnatal folliculogenesis, and not during embryonic oogenesis (Wai *et al.*, 2008).

Amongst these theories, mtDNA CN has been considered to be a major player in heteroplasmy transmission (Jenuth *et al.*, 1996; Cree *et al.*, 2008b), and it is thought that at least one in 200 healthy humans harbour one of 10 pathogenic, potentially disease inducing mitochondrial variants (Elliott *et al.*, 2008; Chinnery *et al.*, 2012). Wallace emphasises the importance of understanding the quantitative mechanisms of mtDNA transmission, and segregation across tissue types during development (Wallace and Chalkia, 2013). Shoubridge and Wai also asked if mtDNA CN is a mechanism that might be able to compensate for these variant's effects when being transmitted (Shoubridge and Wai, 2007).

No study has investigated the degree of mtDNA CN difference between mother and child before, or if mtDNA CN is inherited. Therefore, using a cohort consisting of mother-child pairs I was able to address this.

Aim 5: Identify if mtDNA CN is inherited between healthy mothers and children.

Technical and biological factors introduce differences in mtDNA CN values being reported across the literature, making it difficult to compare results, as well as identify the expected ‘norm’ of mtDNA CN to reference to, and report deviations from that norm. No study has ever investigated mtDNA CN variation in a large European population before, especially in relation to age, gender, and DNA source. Therefore, using the large, European, UKBS cohort data available, I was able to investigate how mtDNA CN changed in relation to these demographic factors. In addition, and perhaps more importantly, from investigating these effects, I am able to identify any covariates that may need to be introduced when studying genome-wide genetic study models, to adjust for factors significantly influencing mtDNA CN changes.

Additionally, this work will highlight any demographic covariates such as gender, age, full blood count data, and heritability that need to be adjusted for in future genome-wide analysis should they be contributing to significant mtDNA CN differences.

3.2 Materials and Methods

3.2.1 Cohort Details and Mitochondrial DNA Copy Number Generation

3.2.1.1 Cohort One - UKBS

Relative mtDNA CN was determined in 3028 (97.96%) UKBS samples using qPCR methods detailed in section 2.8.1. 1512 were males (49.9%) and 1516 females (50.1%). Age was also available for all samples, and full blood count (FBC) data was provided for 2617 UKBS samples. Full details about the UKBS cohort and data available is detailed in chapter 2.

3.2.1.2 Cohort Two – In house 'Cohort Two'

Relative mtDNA CN was determined in 1962 (98.10%) cohort two samples using qPCR methods detailed in section 2.8.1. Age, gender, and blood count data was not available for cohort two samples, however mother-child pairs were known. Full details about the in house 'cohort two' cohort and data available is detailed in chapter 2.

3.2.2 Statistical Analysis

The distribution of raw mtDNA CN in the UKBS and 'cohort two' controls were negatively skewed (Figure 3.1 and Figure 3.2 i's); therefore Z-scoring was applied (Equation 2.3 – section 2.8.1.6) to normalise the mtDNA CN distribution (Figure 3.1 and Figure 3.2 ii's) to perform further parametric statistical tests. All statistical analysis was conducted using IBM SPSS Statistics 22 software (IBM, NY, USA). Independent t-tests (ITT), linear regressions, and Mann- Whitney U (MWU) tests were performed for variable relationships against mtDNA CN. Tested variables included age and gender, as well as FBC data, including absolute and proportional counts of individual cell component data such as basophils, neutrophils, eosinophils, lymphocytes, and monocytes (refer to Table 2.1 for FBC abbreviations). Confidence intervals were accepted at 95%. mtDNA CN, age, and blood cell components were treated as continuous traits, and gender was binary.

3.3 Results

3.3.1 Mitochondrial DNA Copy Number Distributions

To make UKBS and cohort two cohorts comparable for analysis, cohort two was restricted to only include samples whose mtDNA CN was determined from genomic DNA extracted from blood (n=1854, 94.50%) (Table 3.1).

		Mitochondrial DNA Copy-Number Pre-QC (copies/cell)						*Mitochondrial DNA Copy-Number Post-QC (copies/cell)*					
Cohort	**N**	**N**	**Mean**	**Median**	**St. Dev**	**Min**	**Max**	**N**	**Mean**	**Median**	**St. Dev**	**Min**	**Max**
All UKBS (WBC)	3091	3028	178.96	170.20	61.61	38.96	1187.41	3026	178.32	170.17	56.46	38.96	816.15
∟*Males*	1541	1512	180.89	171.12	68.88	38.96	1187.41	1510	179.62	171.08	59.43	38.96	816.15
∟*Females*	1550	1516	177.04	169.84	53.33	58.25	523.22	1516	177.04	169.84	53.33	58.25	523.22
Cohort Two Blood*	1889	1854	1059.66	216.58	14555.96	11.01	583354.66	1735	246.32	208.54	149.38	11.01	888.43
∟Mothers Blood	889	862	1944.20	211.10	21311.93	43.74	583354.66	764	249.70	192.28	170.18	43.74	888.43
∟Children Blood	1000	992	291.04	220.66	532.21	11.01	10884.29	971	243.67	218.84	130.74	11.01	821.62
Mothers Buccal	111	108	1178.81	45.34	5500.83	2.05	47642.69	99	97.96	38.73	146.59	2.05	783.47

*Table 3.1: Breakdown of relative mitochondrial DNA copy number (copies/cell) across UKBS and cohort two samples before and after being filtered for mtDNA CN less than or equal to 900 copies/cell. WBC: DNA extracted from white blood cells. * indicates unknown what cell type DNA was extracted from – refer to Table 2.2.*

Table 3.1 summarises the differences in mtDNA CN between UKBS and cohort two samples before and after a quality control (QC) filter, which only included individuals who had a mtDNA CN of less than or equal to 900 copies/cell, was applied. As shown, the mean and median values for both male and females in UKBS samples are more similar than that of mother, child, or buccal samples in cohort two. mtDNA CN in UKBS samples also ranged from 38.96 to 1187.41 copies/cell, whereas they ranged from 11.01 to 583,354.66 copies/cell in cohort two samples.

To visualise the distributions of mtDNA CN in each cohort, histograms were generated (Figure 3.1). Both raw, and z-scored mtDNA CN were plotted to demonstrate that z-scoring normalised the distributions between cohorts (Figure 3.1-ii's), as raw mitochondrial plots were negatively skewed (Figure 3.1-i's). However, as Table 3.1 shows, the range of mtDNA CN in blood samples from cohort two was extremely high (583343.65 copies/cell) compared to UKBS samples (1148.45 copies/cell). Therefore, to make the cohorts comparable, both cohorts were filtered to only include samples which had a mtDNA CN of less than or equal to 900 copies/cell. Values above 900 copies/cell were considered outliers. This filter removed two (0.06%) UKBS and 119 (6.42%) cohort two individuals respectively. Figure 3.2 shows the frequency distributions of raw mtDNA CN in both cohorts is more normally distributed after being filtered (Figure 3.2-i's), and is similar to z-scored mitochondrial distributions (Figure 3.2-ii's). As a result, both raw, and z-scored mtDNA CN could be used for demographic analyses.

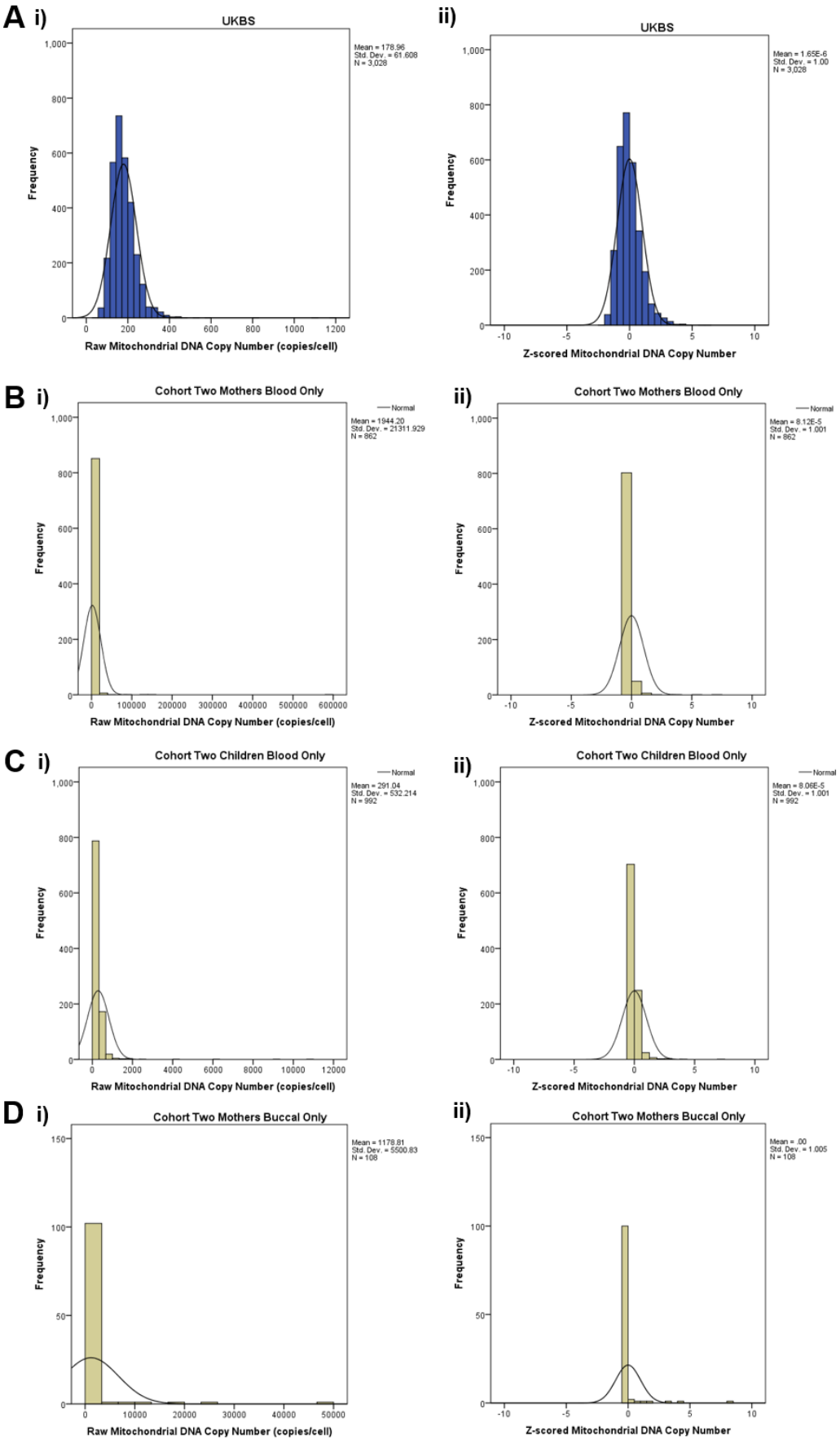


Figure 3.1: Frequency distributions of (i) raw and (ii) Z-scored mitochondrial DNA copy number in unfiltered [A] UKBS (n=3028) and [B] – [D] cohort two samples. [B] Cohort two mothers samples from blood only (n=862), [C] children samples from blood only (n=992), and [D] mothers samples from buccal only (n=108).

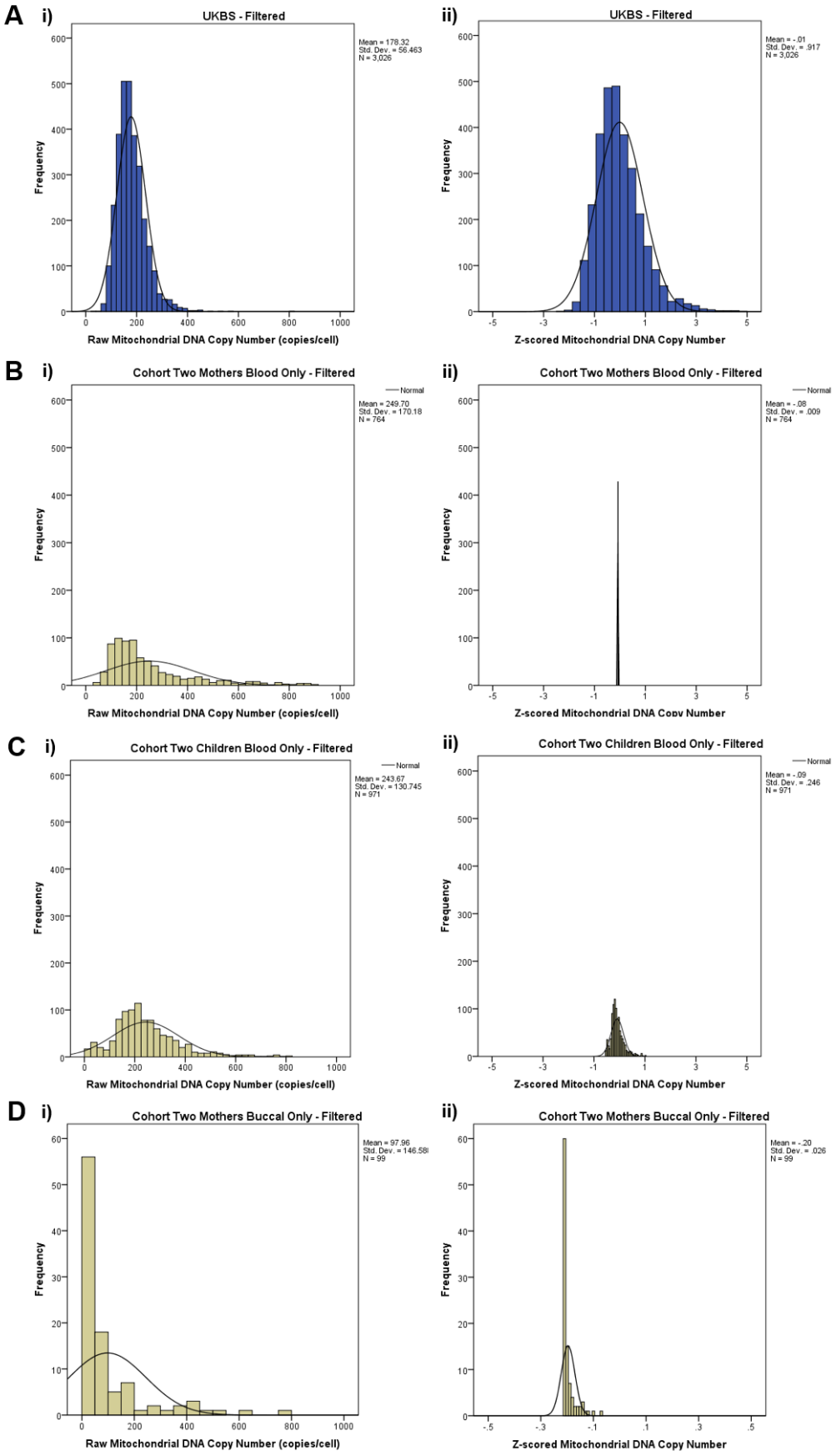


Figure 3.2: Frequency distributions of (i) raw and (ii) Z-scored mitochondrial DNA copy number (mtDNA CN) which have been selected if mtDNA CN is less than or equal to 900 copies/cell. [A] UKBS samples (n=3026) and [B] – [D] cohort two samples. [B] Cohort two mothers samples from blood only (n=764), [C] children samples from blood only (n=971), and [D] mothers samples from buccal only (n=99).

3.3.2 DNA Quality Effect on Mitochondrial DNA Copy Number Determination

As described in section 3.1.2.1, no study has previously used median Ct values to calculate mtDNA CN, because established methods use the mean value of replicates after they have been QC. Using mean Ct values may introduce a bias to mtDNA CN calculations because the mean is finding an average adjusted central point. The impact of this bias across thousands of samples could be detrimental, therefore, using the UKBS cohort, I wanted to compare whether using the median or mean Ct values was most appropriate to calculate mtDNA CN, and identify if there were any significant differences in values across thousands of samples.

Of the 3091 UKBS samples measured, 3088 samples (99.90%) had triplicate Ct measurements generated using qPCR (Figure 3.3). All three replicates were used to calculate median values for each measurement of *B2M*, *MT-ND1,* and *MT-ND4* for each sample. Median derived mtDNA CN was then determined using the same delta Ct calculation used previously (section 2.8.1.5 - Equation 2.2), measuring the proportion of *MT-ND1* or *MT-ND4* to *B2M*.

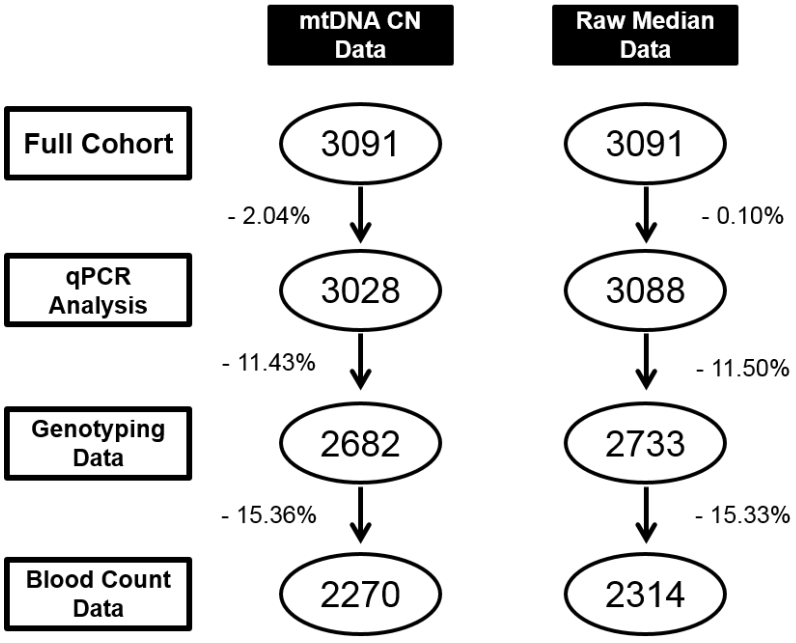
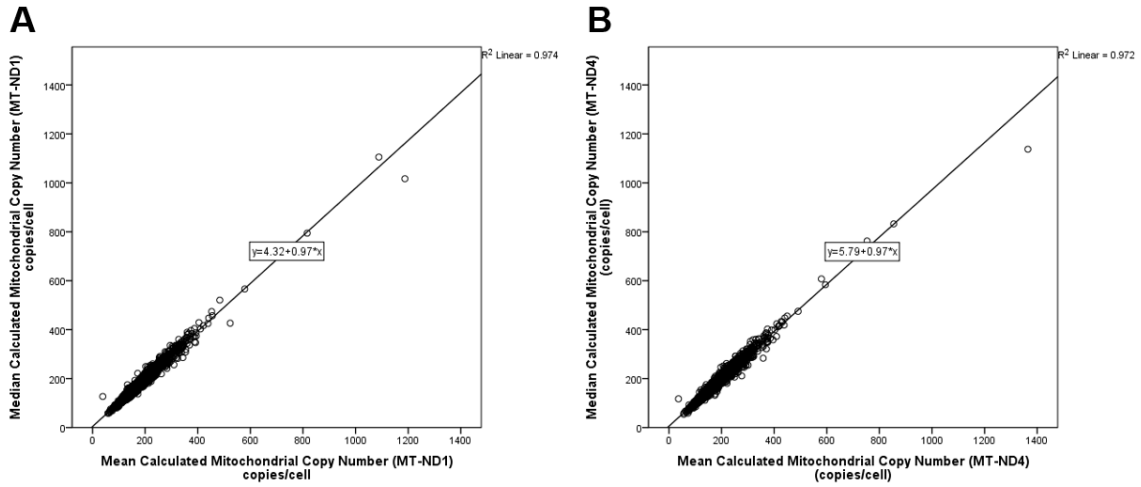


Figure 3.3: Flow diagram summarising the number of UKBS samples with raw Ct data from the qPCR assay, which could either be used to generate median Ct values (requiring a minimum of three replicates), or could be used to calculate mitochondrial DNA copy number (mtDNA CN) (copies/cell) with a minimum of two replicates. Additional arms highlight samples with SNP genotyping and blood count data.

To compare the differences between mean-derived, and median-derived mtDNA CN, linear regression plots were drawn (Figure 3.4A and B), and linear regression analysis and independent t-tests were performed because equal variance was calculated. As indicated in Figure 3.4, there is no significant difference in mean-

derived or median-derived mtDNA CN in UKBS samples, irrespective of using *MT-ND1* (r^2= 0.974, β= 0.987, *p*=0.895) (Figure 3.4A) or *MT-ND4* (r^2= 0.972, β= 0.986, *p*= 1.00) (Figure 3.4B) as the mitochondrial reference gene, but in fact they are strongly correlated.



	Mean (N)	Mean mtDNA CN	Mean mtDNA CN Variance	Median (N)	Median mtDNA CN	Median mtDNA CN Variance	F-test Equal Variance	Two- tailed T-test (Equal Variance) p value
MT-ND1: B2M	3028	178.96	3795.56	3088	178.75	3741.93	0.347	0.895
MT-ND4: B2M	3028	175.93	3772.08	3088	175.93	3698.13	0.292	1.000

Figure 3.4: Relationship between mitochondrial DNA copy number calculated using the mean of Ct values (n=3028), or the median of Ct values (n=3088) in UKBS samples. [A] Mitochondrial DNA copy number calculated using MT-ND1: B2M. [B] Mitochondrial DNA copy number calculated using MT-ND4: B2M. Table summarises the statistics of mitochondrial DNA copy numbers between mean and median options, and the t-tests performed.

Having established there was no difference between mean or median-derived mtDNA CN, I chose to continue using mean-derived mtDNA CN to keep analysis in line with the literature.

Next, I wanted to investigate the quality of the mtDNA by comparing *MT-ND1* derived-values against *MT-ND4* derived values. mtDNA is more likely to develop mutations or deletions than nDNA because mtDNA lacks protective histones and is exposed to harmful ROS products, which damage the DNA (Gemmell *et al.*, 2004; Rygiel *et al.*, 2015). Using UKBS and cohort two samples, linear regression plots were drawn to show the relationship between *MT-ND1* or *MT-ND4* calculated mtDNA CN measurements respectively (Figure 3.5A and Figure 3.6). Additionally, median *MT-ND1* and *MT-D4* values in UKBS samples were compared to investigate if the statistics altered the outcome of the results recorded.

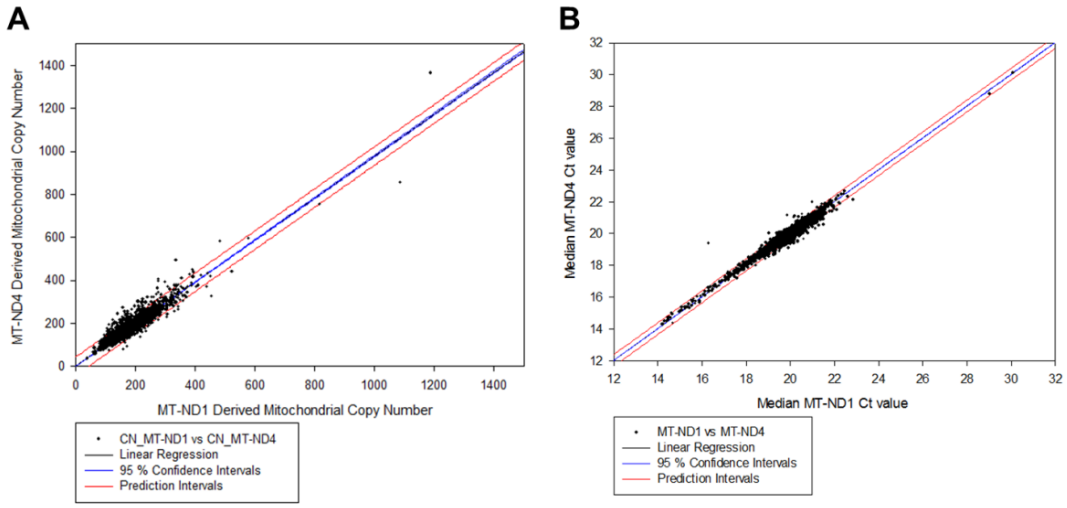


Figure 3.5: Linear regression plot of [A] Mitochondrial DNA copy number derived from using mitochondrial gene MT-ND1 against mitochondrial DNA copy number derived from using mitochondrial gene MT-ND4 in UKBS samples (n=3028), and [B] raw median Ct values generated from triplicate measurements for MT-ND1 plotted against raw median MT-ND4 measurements for each UKBS sample (n=3088). Mitochondrial DNA copy number is copies/cell.

As Figure 3.5A shows, less than 5% of samples sit outside of the prediction interval for *MT-ND1* versus *MT-ND4* calculated mtDNA CN, and were highly correlated (r^2= 0.875, β= 0.935). This was also consistent with the median *MT-ND1* versus *MT-ND4* values, where values strongly correlated (r^2= 0.859, β= 0.927). This demonstrated that *MT-ND1* was proportional to *MT-ND4* in mtDNA from these samples and suggested the mtDNA quality was good.

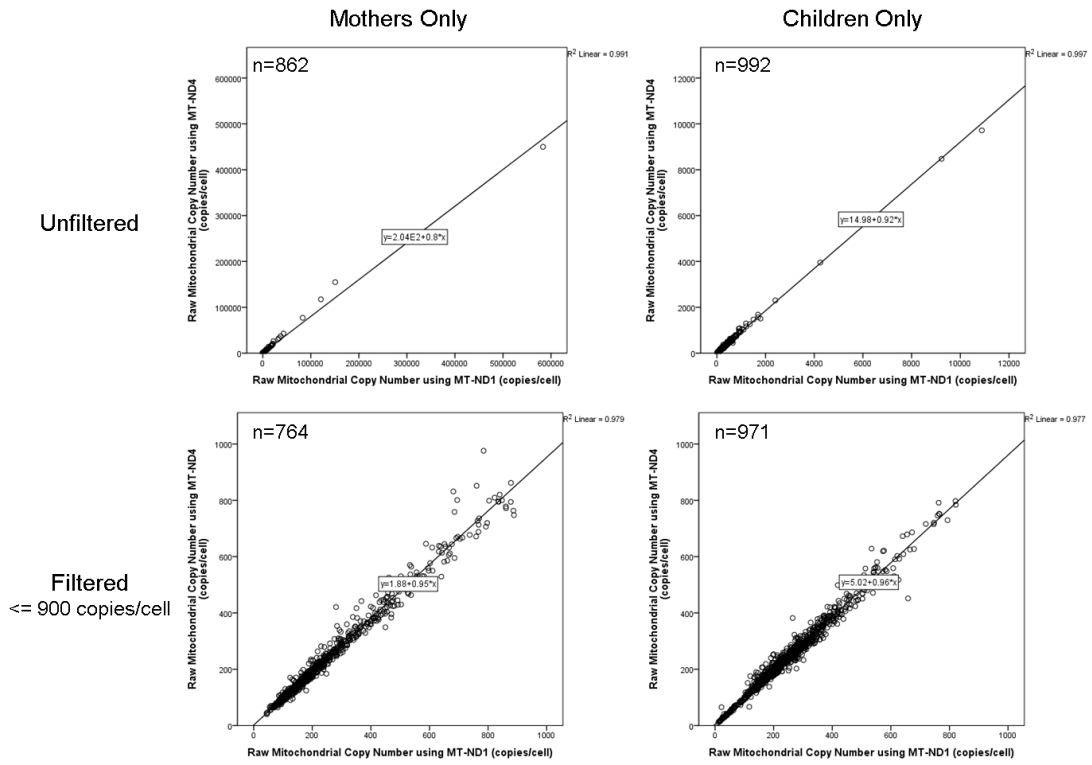


Figure 3.6: Linear scatter plots of MT-ND1 derived mitochondrial DNA copy number (copies/cell) against MT-ND4 derived mitochondrial DNA copy number (copies/cell) in cohort two mothers and children which have or have not been filtered for mitochondrial DNA copy number <=900 copies/cell. Blood extracted DNA samples only.

To investigate mtDNA quality in cohort two DNA samples extracted from blood, *MT-ND1* derived and *MT-ND4* derived mtDNA CN were compared (Figure 3.6). Cohort two was stratified by mothers (n=862) and children (n=992), and then linear regression plots were generated for each. The results showed a positive correlation (unfiltered mothers: r^2=0.991, β=0.995, unfiltered child: r^2=0.997, β=0.998). Additional QC filters were applied, removing individuals with mtDNA CN greater than 900 copies/cell, and the same linear regression analysis was performed. 98 (11.37%) mother samples and 21 (2.12%) children samples were removed during filtering, however the strong positive correlation between *MT-ND1* and *MT-ND4* derived mtDNA CN remained (filtered mothers: r^2=0.979, β=0.989, filtered children: r^2=0.977, β=0.989), thus suggesting cohort two mtDNA samples were good quality.

Following this series of analysis, I had identified that mtDNA quality for cohort two and UKBS samples were of good quality to use for further analysis, and I decided that UKBS and cohort two samples needed to be filtered to only include those with a mtDNA CN of less than or equal to 900 copies/ cell to remove outliers and make cohorts comparable.

3.3.3 Gender Effect on Mitochondrial DNA Copy Number

As detailed in section 3.1.2.2, differences in mitochondrial biogenesis have been reported between genders, however no study has investigated mtDNA CN differences between genders in a large, healthy, European population before. Therefore I was able to utilise the UKBS data to address this.

In line with previous analysis, UKBS samples were filtered to remove outliers, removing two male samples. Subsequently, samples with mtDNA CN values less than or equal to 900 copies/cell were included (n=3026). 1510 males (49.9%) and 1516 females (50.1%) passed QC and were used for analysis. Individuals were stratified by gender, and had boxplots generated to compare both raw, and z-scored mtDNA CN between genders (Figure 3.7).

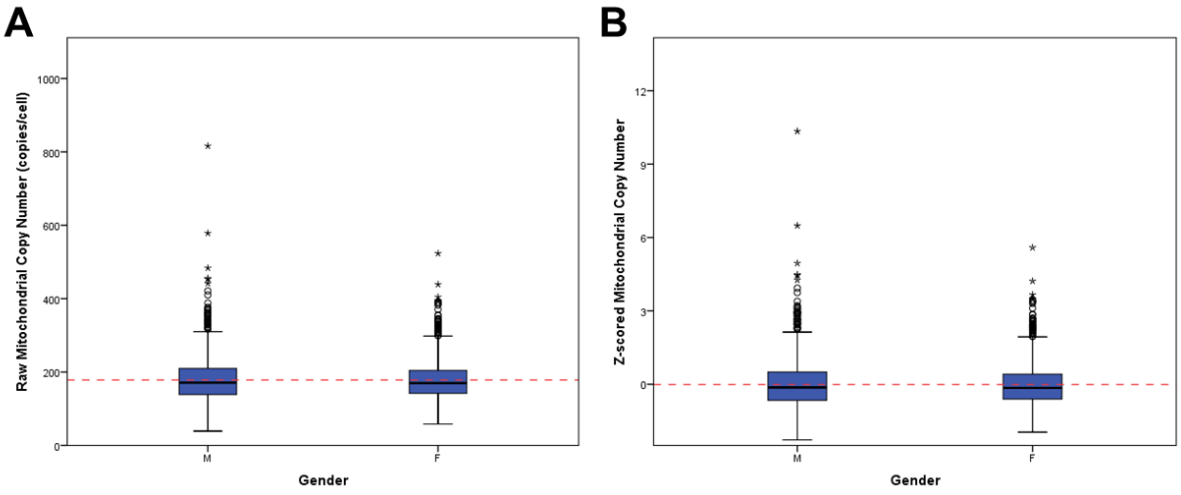


Figure 3.7: Boxplots of [A] raw and [B] Z-scored mitochondrial DNA copy number differences between males (n=1510) and females (n=1516) in the UKBS samples. Red dotted lines indicate population means (raw: μ=178.32) ±56.46 s.d).

The average mtDNA CN in the UKBS cohort was 178.32 copies/cell ±56.46 s.d. Independent T-tests (ITT) and Mann Whitney-U (MWU) tests performed on both raw, and Z-scored mtDNA CN data reported no significant difference in mtDNA CN between males (μ= 179.62 copies/cell ±59.43 s.d) and females (μ= 177.04 copies/cell ±53.33 s.d) (ITT: p= 0.209, MWU: p= 0.550) (Figure 3.7). This suggests that gender does not affect mtDNA CN in the general population.

3.3.4 Age Effect on Mitochondrial DNA Copy Number

Across the literature, mtDNA CN has been reported to change with age, but these are only in small populations of individuals. No study has investigated the age related effects on mtDNA CN changes in large populations of individuals; therefore, using the UKBS data, I investigated the extent mtDNA CN changes in relation to age and gender.

To ensure UKBS samples were suitable to perform age-associated analysis, frequency distribution plots were generated to summarise the ages of all UKBS individuals (Figure 3.9A), and then by females and males only (Figure 3.9B and C). Age distributions were normal and were therefore suitable for testing. Details of age distributions are in Table 3.2.

To keep analysis consistent, UKBS samples with mtDNA CN less than or equal to 900 copies/cell (n=3026) were analysed against each individual's age, and results were then stratified by gender. Linear regression tests were performed, and linear plots were drawn to identify the relationship with mtDNA CN and age (Figure 3.8).

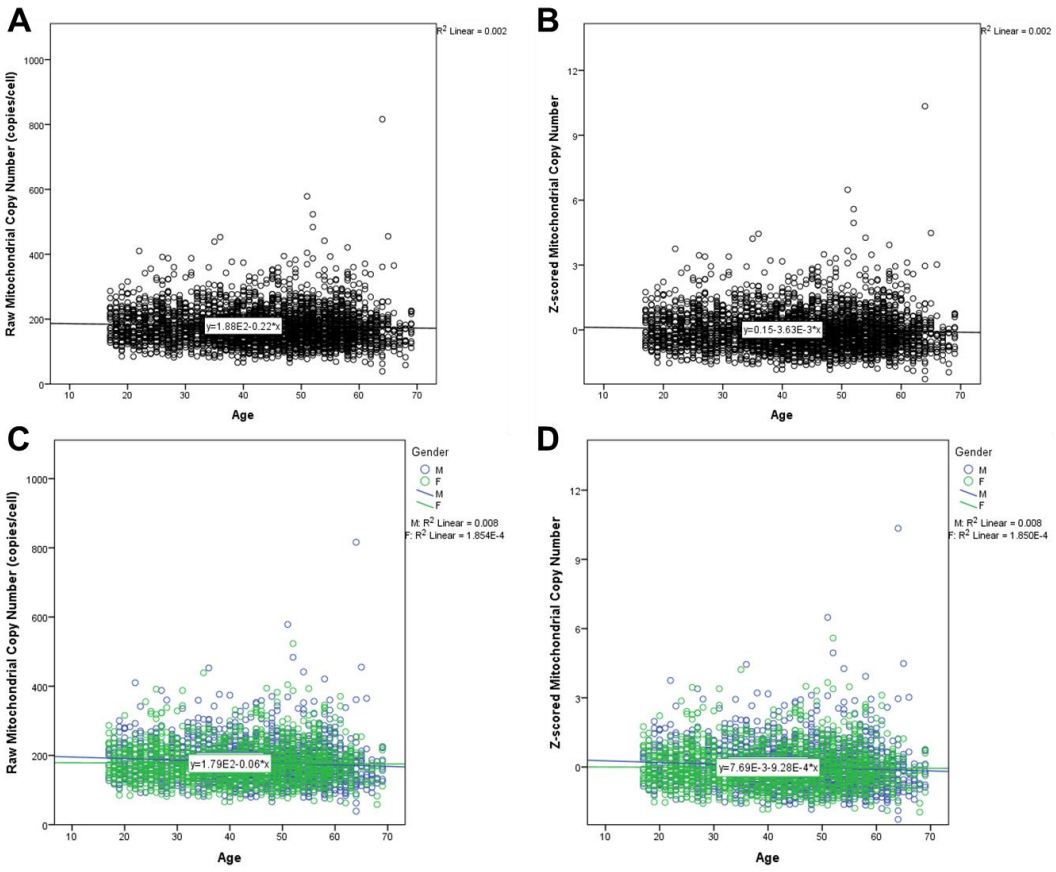


Figure 3.8: Linear regression plots of [A] and [C] raw or [B] and [D] Z-scored mitochondrial DNA copy number changes with age in UKBS samples. [A] and [B] depict all UKBS samples (n=3026) and [C] and [D] depict males (n=1510, blue) and females (n=1516, green).

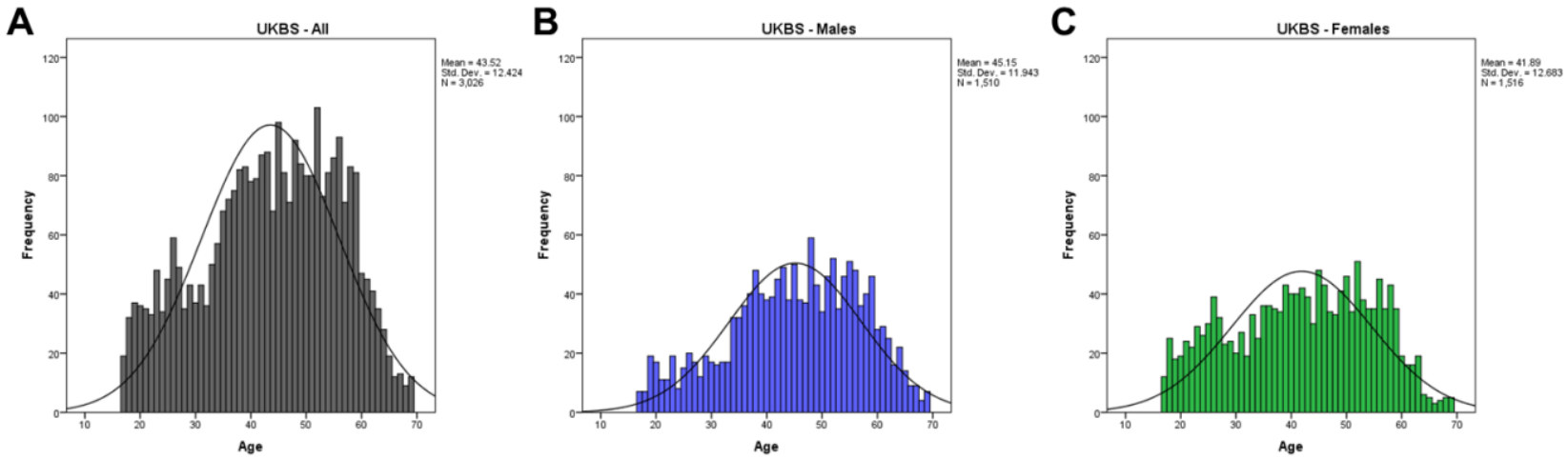


Figure 3.9: Frequency distribution curves of ages in UKBS samples (n=3026), stratified by gender (n=1510 males, n=1516 females). Black line indicates normal distribution curve.

A summary of the results generated from linear regression tests are presented in Table 3.2. Across the UKBS cohort, mtDNA CN statistically significantly decreases with age (p= 0.007, β= -0.049, r^2= 0.002) however, when stratified by gender, there was no statistically significant decrease in mtDNA CN with age in females (p= 0.596, β= -0.014, r^2= 1.85E-04), but there was a statistically significant decrease in mtDNA CN with age in males (p= 4.42E-04, β= -0.090, r^2= 0.008). Therefore the significant decrease reported in the whole population was a result of a gender specific effect.

These results show that in the general, healthy population, mtDNA CN does statistically significantly change with age in males, but not in females, and this may give reason to why males are more affected by age-related, and mitochondrial diseases than females (Hudson *et al.*, 2007a; Frank, 2012; Hudson *et al.*, 2013a). However, these results are likely not biologically significant because for every 10 years mtDNA CN decreases by 0.11 copies/cell, 0.18 copies/cell, and 0.03 copies/cell in the entire UKBS cohort, in UKBS males only, and in UKBS females only respectively.

UKBS Cohort	Number (N)	Age Range (years)	Median Age (± standard deviation)	Mean Copy Number (copies/cell)	Linear Relationship of Age vs Copy Number		
					R^2	β	*p*
All	3026	17-69	45 (12.42)	178.32	0.002	-0.049	0.007
Males	1510	17-69	46 (11.94)	179.62	0.008	-0.090	4.42e-04
Females	1516	17-69	43 (12.68)	177.04	1.85e-4	-0.014	0.596

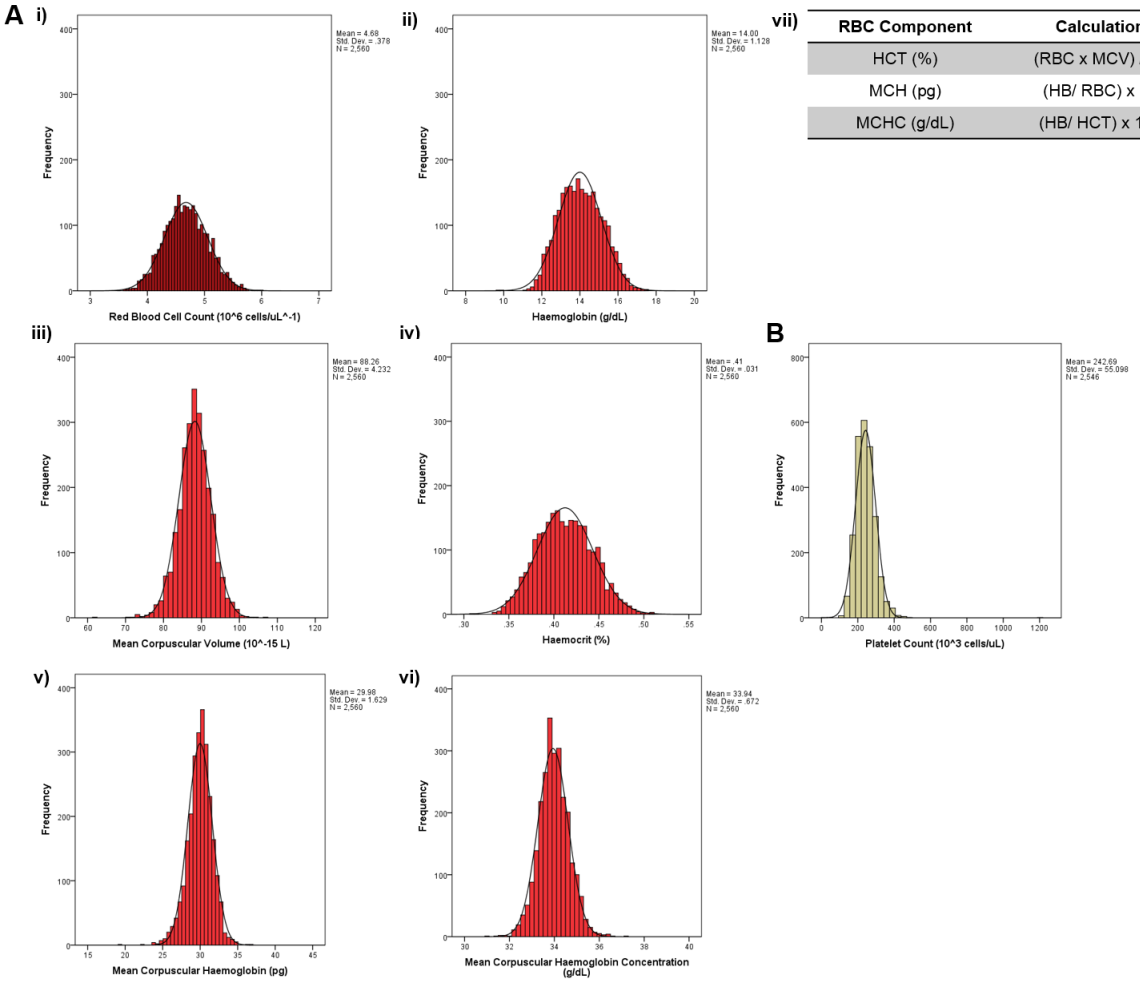
Table 3.2: Summary of UKBS cohort ages, mean mitochondrial DNA copy number, and linear relationship statistics of mitochondrial DNA copy number changes with age, and then stratified by gender.

3.3.5 Blood Composition Effect on Mitochondrial DNA Copy Number

Mitochondrial biogenesis has been reported to change in blood cells depending on individual blood cellular environment, as well as overall blood cellular composition (detailed in section 3.1.2.4). Utilising FBC data available for UKBS samples, I was able to investigate mtDNA CN differences in blood across a large, healthy, European cohort.

Prior to performing any mtDNA CN analysis, histograms of each individual blood cell component were generated, but only for individuals who had a mtDNA CN of less than or equal to 900 copies/cell (n= 3026), to keep analysis consistent (Figure 3.10 and Figure 3.11). As a result, two individuals with blood count data were excluded. 2560 individuals (82.82%) had blood count and mtDNA CN data, of which 1267 (49.49%) were males, and 1293 (50.51%) were females. As Figure 3.10 and Figure 3.11 show, the distributions of all FBC subcomponents were normal for both counts, and overall proportions.

Histogram frequencies of each blood component for UKBS samples were compared to the normal ranges expected for full differential blood count information (Vejabhuti Curry, 2015; Mayo Clinic Staff, 2016) to ensure samples were representative of healthy individuals in populations (Table 3.3). As Table 3.3 shows, all FBC subcomponents were in the expected normal ranges, even when stratified by gender. Basophil counts and proportions were the only subcomponents that were not in the expected ranges, but this may be a result of 2337 samples having their value assigned to zero instead of using the verbose text data which was inconclusive to analysis (section 2.3 - Table 2.1). Normal FBC data ranges reassured that further investigations of mtDNA CN changes in each blood cell type would be representative of active mtDNA CN changes in large populations of healthy, European people, and would not be an artefact of abnormal whole blood cellular proportions contributing to changeable and incomparable mtDNA CN levels.



RBC Component	Calculation
HCT (%)	(RBC x MCV) / 10
MCH (pg)	(HB/ RBC) x 10
MCHC (g/dL)	(HB/ HCT) x 100

Figure 3.10: Frequency distribution plots of blood cell components in UKBS samples that have mitochondrial DNA copy number less than or equal to 900 copies/cell (n=3026). [A] Graphs related to red blood cell counts (n=2560, red); (i) Absolute blood cell counts (RBC), (ii) haemoglobin (HB), (iii) mean corpuscular volume (MCV), (iv) haemocrit (HCT, (v) mean corpuscular haemoglobin (MCH), and (v) mean corpuscular haemoglobin concentration (MCHC). [Avii] Table explaining how HCT, MCH, and MCHC data are determined from RBC count data. [B] Absolute count of platelets (n=2546, yellow). Curves outline normal distributions.

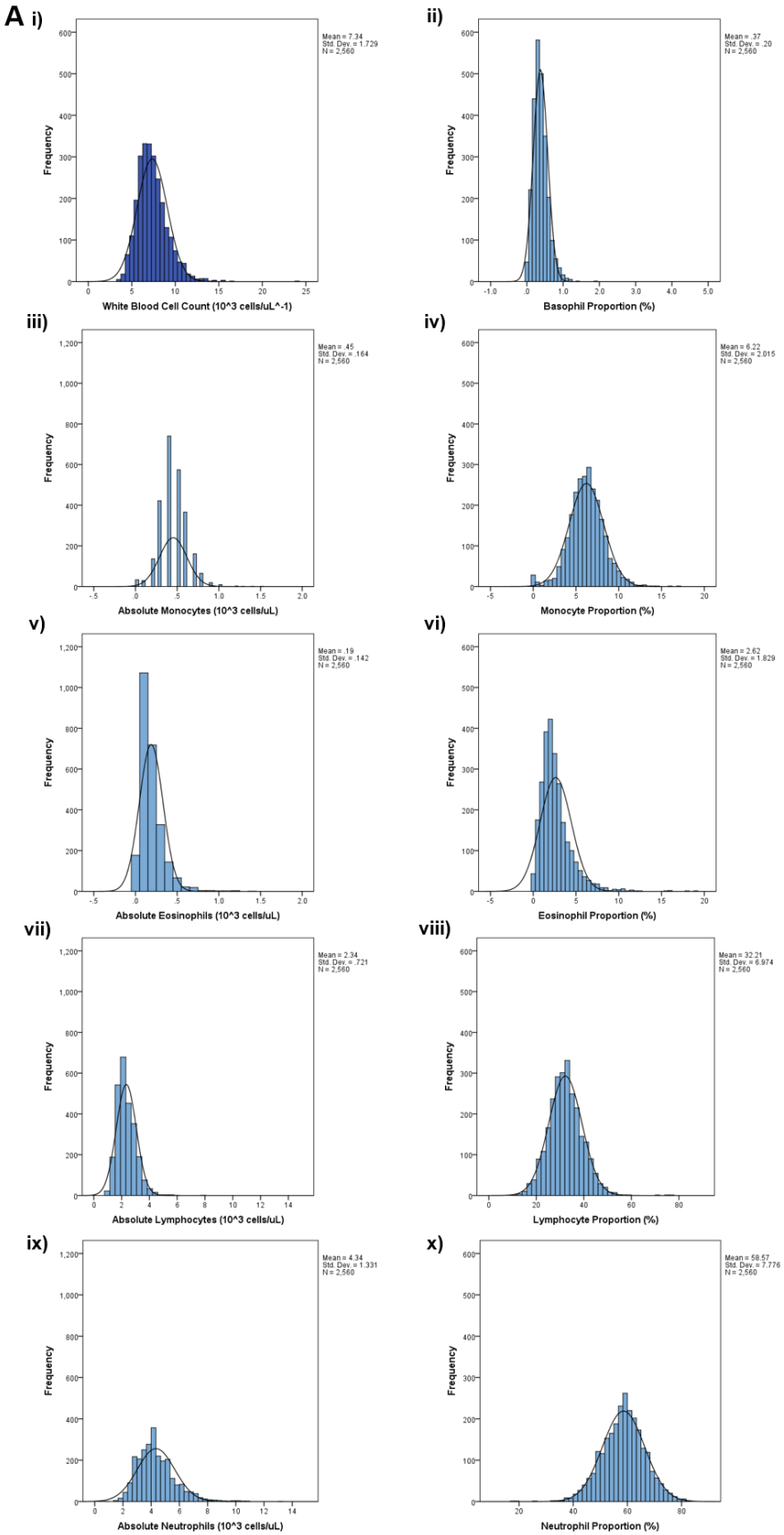


Figure 3.11: Frequency distribution plots of white blood cell cell components (n=2560) in UKBS samples that have mitochondrial DNA copy number less than or equal to 900 copies/cell (n=3026). [i, iii, v, vii, ix] Absolute counts. [ii, iv, vi, viii, x] Relative proportions. (ii) Basophil proportions (BP). (iii and iv) Monocytes (MA and MP). (v and vi) Eosinophils (EA and EP). (vii and viii) Lymphocytes (LA and LP). (ix and x) Neutrophils (NA and NP). Absolute basophil data not shown because data is binary. Curves outline normal distributions.

			All UKBS		UKBS Males*		UKBS Females*		* ITT	* MWU
Blood Component	**Units**	**Normal Range**	**N**	**Mean (+/- SD)**	**N**	**Mean (+/- SD)**	**N**	**Mean (+/- SD)**	***p***	***p***
Platelets	10^3 cells/μL	150 - 450	2546	242.69 (55.10)	1258	227.66 (45.88)	1288	257.37 (59.25)	8.32E-44	<5.00E-04
Red Blood Cells	10^6 cells/ μL^-1	3.90 - 5.03	2560	4.68 (0.38)	1267	4.90 (0.31)	1293	4.46 (0.30)	9.77E-233	<5.00E-04
Haemoglobin	g/dL	12.0 - 15.5	2560	14.00 (1.13)	1267	14.77 (0.85)	1293	13.24 (0.80)	0.00E+00	<5.00E-04
Haemocrit	%	0.349 - 0.500	2560	0.41 (0.03)	1267	0.43 (0.02)	1293	0.39 (0.02)	0.00E+00	<5.00E-04
Mean Corpuscular Haemoglobin	pg	27 - 33	2560	29.98 (1.63)	1267	30.20 (1.47)	1293	29.76 (1.75)	3.92E-12	<5.00E-04
Mean Corpuscular Haemoglobin Concentration	g/dL	33 - 36	2560	33.94 (0.67)	1267	34.15 (0.65)	1293	33.74 (0.63)	5.05E-56	<5.00E-04
Mean Corpuscular Volume	10^-15 L	80 - 96	2560	88.26 (4.23)	1267	88.39 (3.90)	1293	88.13 (4.53)	1.23E-01	0.318
White Blood Cells	10^3 cells/ μL^-1	3.5 - 10.5	2560	7.34 (1.73)	1267	7.12 (1.64)	1293	7.56 (1.79)	9.19E-11	<5.00E-04
Absolute Basophils	10^3 cells/μL	0.02 - 0.1	2560	0.01 (0.03)	1267	0.01 (0.03)	1293	0.01 (0.03)	4.20E-02	
Basophil Proportion	%	< 1 - 2%	2560	0.37 (0.20)	1267	0.38 (0.20)	1293	0.37 (0.20)	7.90E-02	0.054
Absolute Eosinophils	10^3 cells/μL	0.02 - 0.5	2560	0.19 (0.14)	1267	0.20 (0.14)	1293	0.18 (0.14)	6.00E-03	<5.00E-04
Eosinophil Proportion	%	1 - 6%	2560	2.62 (1.83)	1267	2.83 (1.83)	1293	2.43 (1.81)	3.33E-08	<5.00E-04
Abolute Monocytes	10^3 cells/μL	0.2 - 1.0	2560	0.45 (0.16)	1267	0.47 (0.16)	1293	0.43 (0.16)	3.69E-11	<5.00E-04
Monocyte Proportion	%	2 - 10%	2560	6.22 (2.02)	1267	6.68 (2.03)	1293	5.77 (1.90)	3.76E-31	<5.00E-04
Absolute Lymphocytes	10^3 cells/μL	1.0 - 3.0	2560	2.34 (0.72)	1267	2.23 (0.78)	1293	2.45 (0.64)	3.27E-15	<5.00E-04
Lymphocyte Proportion	%	20 - 40%	2560	32.21 (6.97)	1267	31.51 (7.02)	1293	32.89 (6.86)	5.46E-07	<5.00E-04
Absolute Neutrophils	10^3 cells/μL	2.0 - 7.0	2560	4.34 (1.33)	1267	4.20 (1.23)	1293	4.48 (1.41)	2.06E-07	<5.00E-04
Neutrophil Proportion	%	40 - 80%	2560	58.57 (7.78)	1267	58.60 (7.90)	1293	58.55 (7.65)	8.63E-01	0.995

*Table 3.3: Summaries of the means and standard deviations (s.d) of each blood cell component generated from full blood count data in UKBS samples (n=3026), and stratified by gender (n=1510 males and n=1516 females). Recorded means are compared to the expected normal blood count ranges (Vejabhuti Curry, 2015; Mayo Clinic Staff, 2016). * indicates tests performed between male and female blood component means. ITT: Independent T-test. MWU: Mann-Whitney U test.*

To identify blood composition differences between genders in these samples (Table 3.3), UKBS samples were stratified by gender, and then had mean values generated for each FBC subcomponent. Boxplots were generated to visualise the difference in means for each cellular component between genders (Figure 3.12 and Figure 3.13). ITT and MWU tests were then performed on the means of each FBC subcomponent between females and males, and p-values were recorded in Table 3.3. As Table 3.3 shows, there are significant differences in all FBC subcomponents between males and females, except in MCV, BA and BP, and NP. Of the nucleated cells, PLT, MP, and LA had the strongest significant differences between males and females (ITT: p= 8.32E-44, p= 3.76E-31, and p= 3.27E-15 respectively).

UKBS DNA samples were extracted from WBC pellets, as they contain the most genomic DNA in blood. Therefore, to investigate mtDNA CN changes in blood components in healthy populations, UKBS FBC data was restricted to only include WBCs. Scatter plots were generated to visualise the relationship between mtDNA CN and each WBC component – basophils, eosinophils, monocytes, lymphocytes, and neutrophils, as well as platelet counts (Figure 3.14). Both absolute counts and proportions were compared to identify if results were similar, and linear regression analysis was conducted to identify any significant relationships (Table 3.4). Linear regression was not performed on BA because the variable was binary. RBCs do not contain DNA, therefore they were not included in any mtDNA CN analysis, and neither were the related HB, HCT, MCH, MCHC, and MCV components.

As Figure 3.14 and Table 3.4 show, there were changes in mtDNA CN relative to each WBC subcomponent absolute amount and WBC proportion, except MA and PLT. Overall, WBC had a significant decrease in mtDNA CN as cell count increased (p= 7.02E-09), but when stratified by subcomponents, this effect was a result of a strong significant decrease in mtDNA CN with cell count in NA and NP (p= 8.84E-22 and p= 3.59E-27 respectively). Other cellular subcomponents reported a significant increase in mtDNA CN as cell count increased. EA, LA, BP, MP, EP, and LP all had p values less than 0.05.

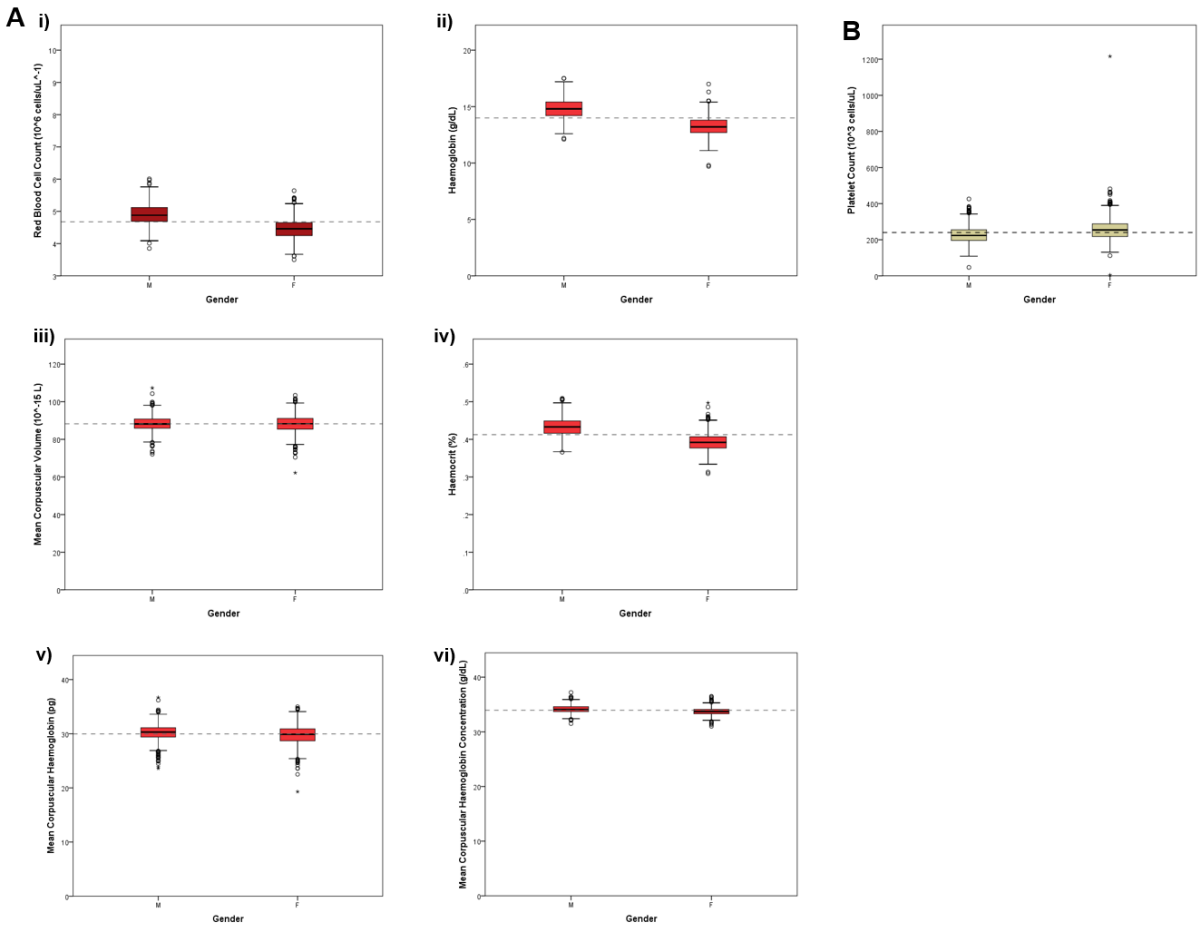


Figure 3.12: Boxplots demonstrating the difference in blood cell compositions between males (M, n=1267) and females (F, n=1293) in 2560 UKBS individuals which had full blood count determined. [A] Boxplots related to red blood cell count data; (i) red blood cell counts (RBC), (ii) haemoglobin (HB), (iii) mean corpuscular volume (MCV), (iv) haemocrit (HCT), (v) mean corpuscular haemoglobin (MCH), and (v) mean corpuscular haemoglobin concentration (MCHC). [B] Absolute count of platelets. Dotted line indicates population mean. Statistics are in Table 3.3.

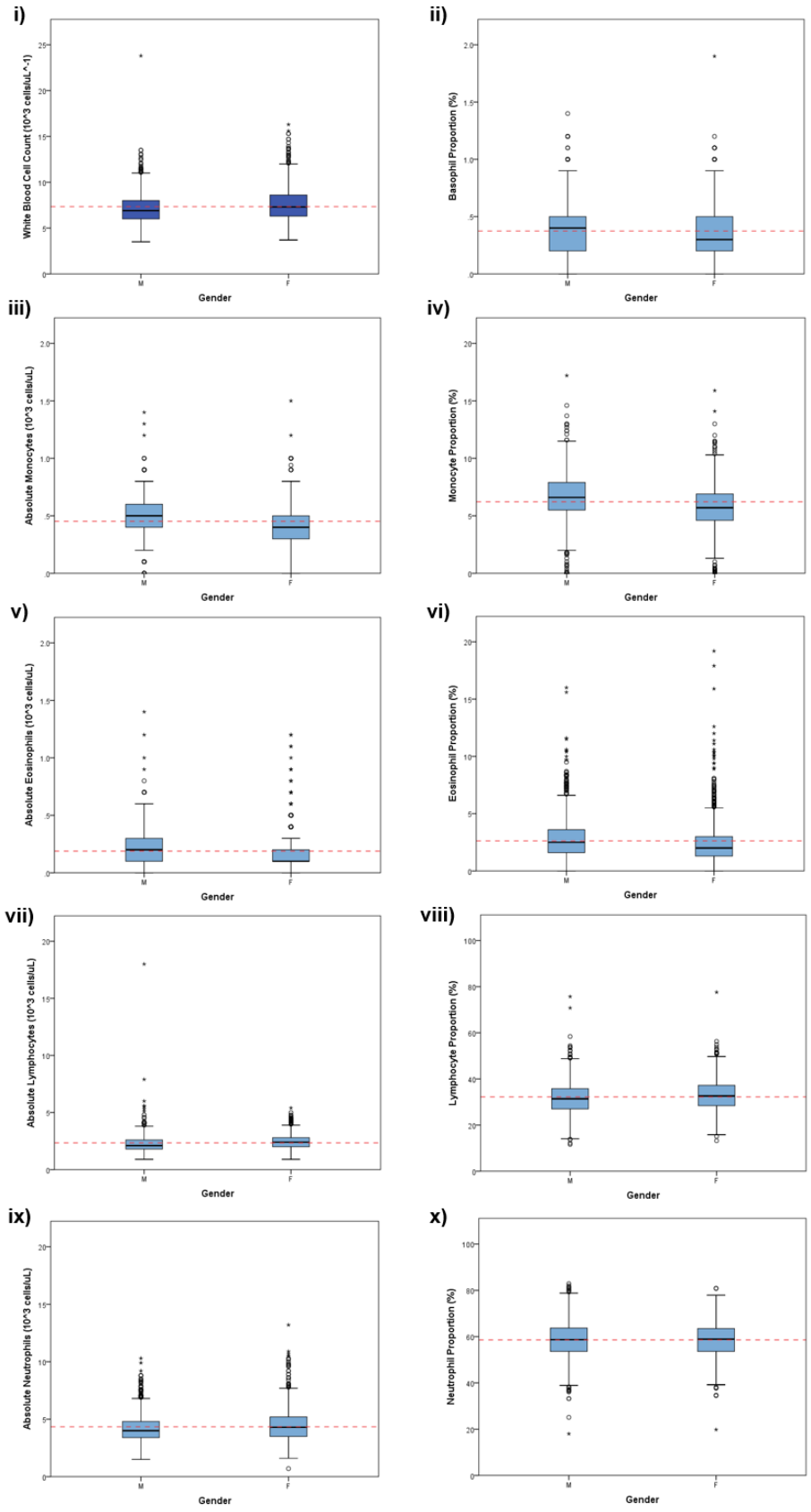


Figure 3.13: Boxplots demonstrating the difference in white blood cell compositions between males (M, n=1267) and females (F, n=1293) in 2560 UKBS individuals which had full blood count determined. [i, iii, v, vii, ix] Absolute counts. [ii, iv, vi, viii, x] Relative proportions. (ii) Basophil proportions (BP). (iii and iv) Monocytes (MA and MP). (v and vi) Eosinophils (EA and EP). (vii and viii) Lymphocytes (LA and LP). (ix and x) Neutrophils (NA and NP). Absolute basophil data not shown because data is binary. Dotted line indicates population mean. Statistics are in Table 3.3.

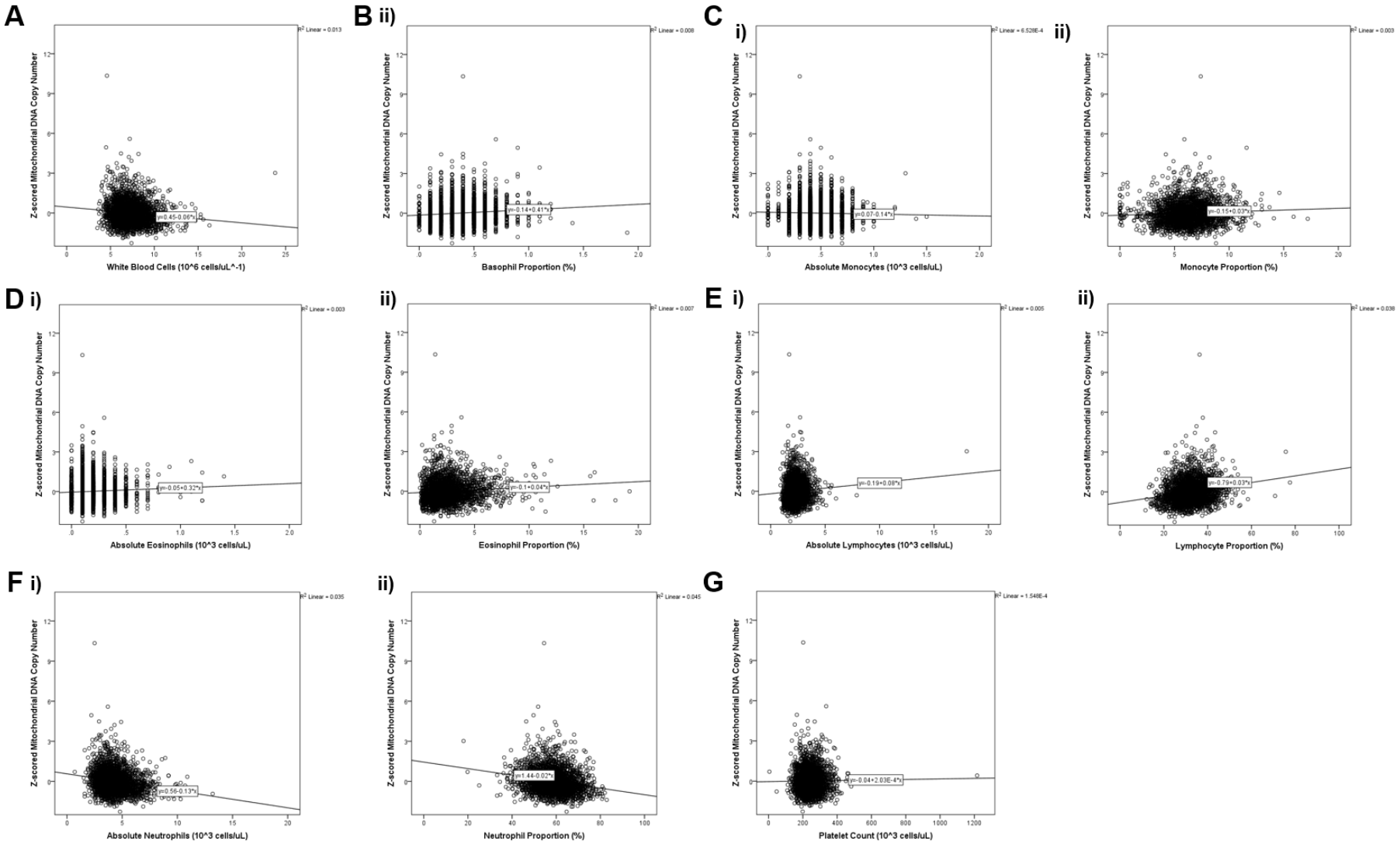


Figure 3.14: Scatter plots showing the relationships between z-scored mitochondrial DNA copy number and [A] Full white blood cell components in UKBS samples – stratified by [B] basophils, [C] monocytes, [D] eosinophils, [E] lymphocytes and [F] neutrophils, as well as [G] platelet counts. (i) Indicates absolute count data. (ii) Indicates proportion data. Table 3.4 summarises the linear regression analysis results for each respective graph.

Mitochondrial DNA Copy Number in European Populations

		Linear Regression Z-scored mtDNA CN		
Graph	**Blood Component**	**R^2**	**Beta**	**P**
A	WBC	0.013	-0.114	7.02E-09
Bii	BP	0.008	0.091	4.00E-06
Ci	MA	0.001	-0.026	1.96E-01
Cii	MP	0.003	0.058	3.00E-03
Di	EA	0.003	0.051	1.00E-02
Dii	EP	0.007	0.083	2.40E-05
Ei	LA	0.005	0.068	6.07E-04
Eii	LP	0.038	0.194	3.89E-23
Fi	NA	0.035	-0.188	8.84E-22
Fii	NP	0.045	-0.211	3.59E-27
G	PLT	1.55E-04	0.012	5.30E-01

Table 3.4: Summarises of linear regression analysis conducted to identify the relationship between Z-scored mitochondrial DNA copy number and absolute cell counts or total cell proportions of individual white blood cell types in UKBS samples. Table correlates to graphs in Figure 3.14.

3.3.6 Inheritance of Mitochondrial DNA Copy Number

mtDNA is maternally inherited, however the inheritance of mtDNA CN is not fully defined. To investigate this, mtDNA CN was measured in cohort two mother-child pairs. Of the 2000 cohort two samples, 855 mother-child pairs had mtDNA CN calculated from DNA extracted from blood. To keep analysis consistent, both mother and child samples were filtered to include mtDNA CN less than or equal to 900 copies/cell. This removed 112 pairs (13.10%), leaving 743 pairs (86.90%) to analyse. Samples were restricted to mother and child samples only from the pairs, and then frequency distribution plots of mtDNA CN were generated to compare frequencies (Figure 3.16).

As Figure 3.16 shows, there is a negative skew in raw mtDNA CN in cohort two mothers and child pairs. Mothers have a larger range in unfiltered mtDNA CN compared to children. Z-scoring both unfiltered, and filtered mtDNA CN normalised the distributions and made them comparable.

Linear regression analysis was performed on both unfiltered and filtered mother-child pairs to ensure that the filtering process had not introduced a bias to the results observed in the unfiltered pairs. Unfiltered mother-child pairs (n=855 pairs) showed there was no significant relationship between mother and child mtDNA CN (r^2= 1.18E-04, β= -0.011, *p*= 0.751) (Figure 3.15A and B). A similar result was also reported in filtered mother-child pairs (n=743 pairs) (r^2= 0.003, β= 0.055, *p*= 0.134) (Figure 3.15C and D). This data is not able to detect heritability, and does not support the hypothesis that mtDNA CN is inherited from mother to child, or that mtDNA CN is controlled only by genetics.

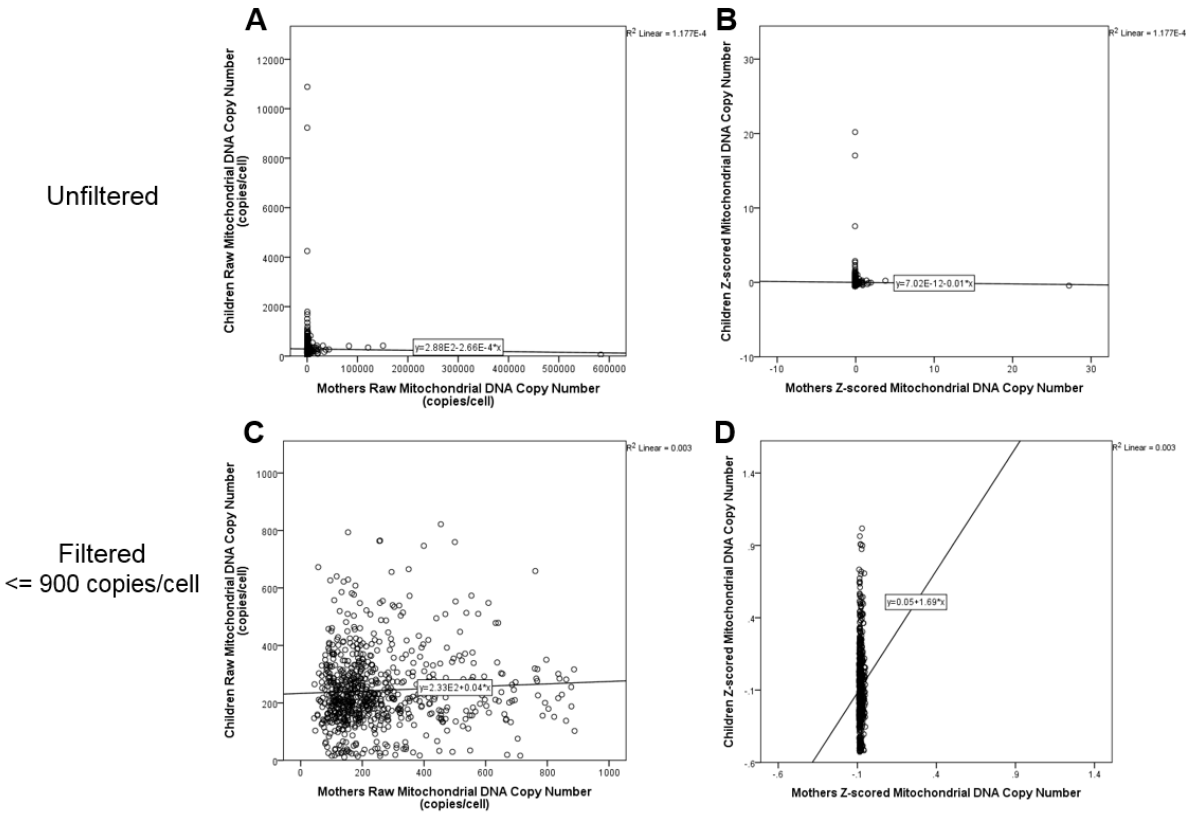


Figure 3.15: Scatter plots of mother versus child [A] and [C] raw mitochondrial DNA copy number or [B] and [D] z-scored mitochondrial DNA copy number to show if mitochondrial DNA copy number is inherited from mother to child. [A] and [B] Unfiltered mitochondrial DNA copy number (n=855 pairs). [C] and [D] Filtered for mitochondrial DNA copy number <= 900 copies/cell in mother and child (n=743 pairs).

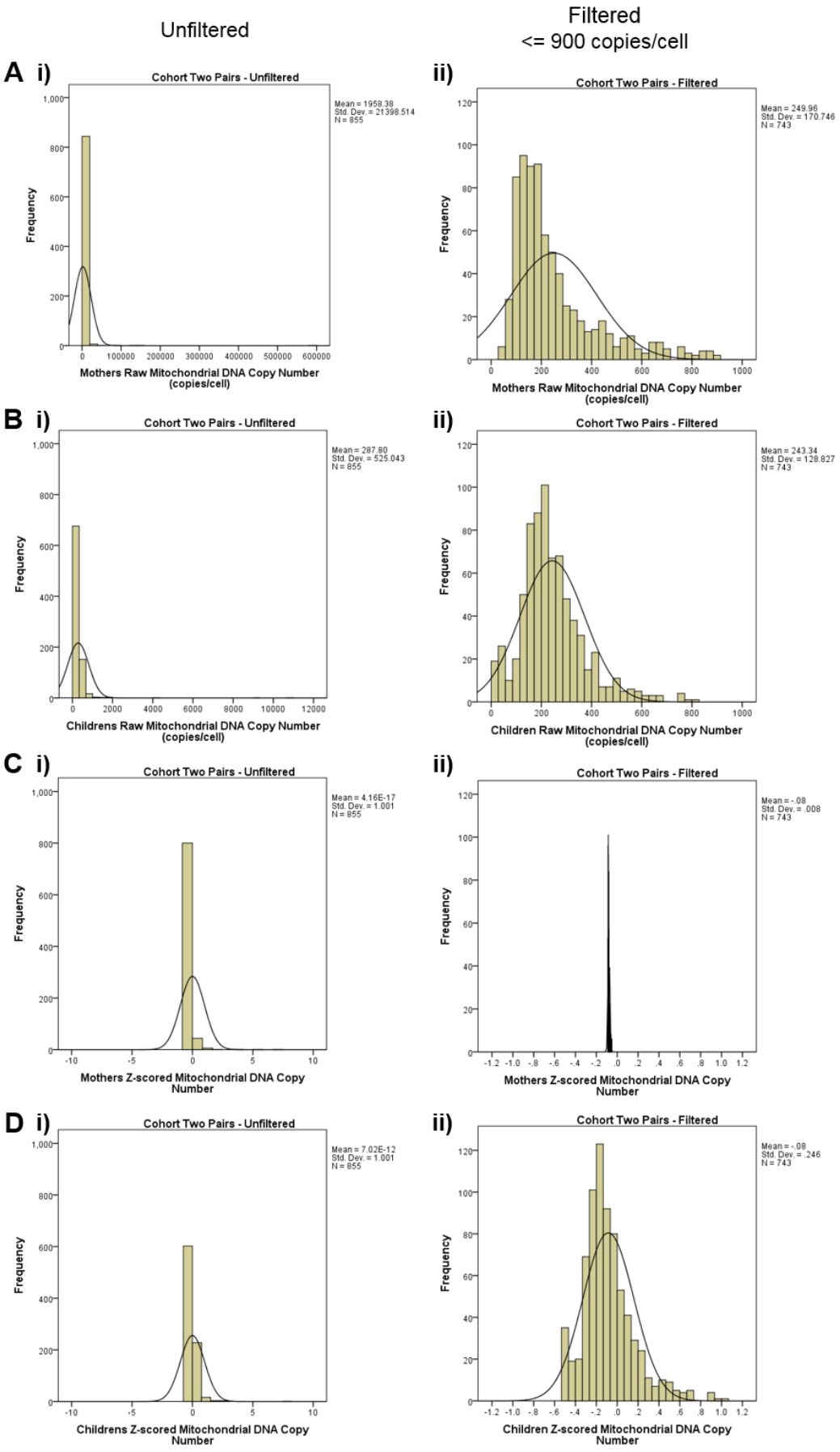


Figure 3.16: Frequency distributions of [A] and [B] raw and [C] and [D] Z-scored mitochondrial DNA copy number in cohort two mother-child pairs which have (i's) or have not (ii's) been filtered to include only mitochondrial DNA copy number less than or equal to 900 copies/cell.

3.4 Discussion

Current literature is limited by power when reporting active mtDNA CN changes in the population, especially in relation to disease or across tissue types, because sample sizes are usually in the tens or hundreds and is not representative of large populations. This has resulted in conflicting results being reported about mtDNA CN changes, especially in disease (reviewed by (Malik and Czajka, 2013). My study is the first to address this limitation, as my results show the natural variation in mtDNA CN in a large, European population relevative to age, gender, and natural blood composition. In the results described, I have identified that there is no significant difference in relative mtDNA CN calculated using the mean or median of Ct values. I also report no significant mtDNA CN difference in males and females, however there is a statistically significant decrease in mtDNA CN with age in males only, albeit this may not be biologically significant because male's mtDNA CN decreases by 0.18 copies/cell every 10 years. Blood cellular composition is seen to affect mtDNA CN changes – having a positive effect in monocytes, eosinophils, lymphocytes, and basophils, but having a negative effect in neutrophils. What's more is I did not report any correlation between mtDNA CN between mother and child in an independent cohort, suggesting mtDNA CN is not inherited.

Cohort two samples were restricted to blood-extracted DNA only because tissue source affects DNA yield and quality, and buccal swabs are known to produce low yield and degrade quickly (Livy *et al.*, 2012; Ghatak *et al.*, 2013). Therefore, to make UKBS and cohort two samples comparable, DNA needed to be from the same source.

The differences in mtDNA CN ranges, originally reported between cohort two and UKBS samples before they were filtered, may have been as a result of a combination of DNA quality and qPCR limitations, whereby cohort two appeared to have poorer DNA quality than UKBS samples. When I was measuring cohort two samples using qPCR, it was more difficult to generate consistent replicates for each sample, and *B2M* Ct values were far more variable than for UKBS samples. Cohort two samples were supplied to the lab in a lower volume (10 µl) rather than the 200 µl volume received for the UKBS samples, and cohort two samples had been kept at -80°C for long-term storage prior to analysis. The lower volume makes the samples more susceptible to freeze-thaw damage (Ross *et al.*, 1990), and as a result DNA damage could have decreased qPCR amplification efficiency, giving rise to more varied qPCR

Ct readings (Ponti *et al.*, 1991; Gonzalez-Hunt *et al.*, 2016). nDNA is also much harder to extract from low quality samples than mtDNA, because nDNA is present in much lower copies. If cohort two samples were poor quality samples before DNA extraction methods were applied, this might explain why the *B2M* Ct value variation reported would have been so inconsistent. This cumulative effect would impact the overall mtDNA CN generated for each sample across the cohort, and could have contributed to the large mtDNA CN ranges observed. Rohland and colleagues suggest that if handling DNA from poor quality samples, using a silica column-based method optimises both yield and DNA quality (Rohland *et al.*, 2010), which might improve future DNA management.

Irrespective of the mtDNA CN differences reported, the overall distribution between the cohorts was similar. The strong, positive correlation between *MT-ND1*-derived and *MT-ND4*-derived mtDNA CN suggests good quality mtDNA, however there is a shift in mtDNA CN towards *MT-ND4* in Figure 3.5A. Although, looking at raw Ct values in Figure 3.5B, *MT-ND1* and *MT-ND4* are more tightly correlated, which suggests that the real-time assay is more selective towards *MT-ND1: B2M* conditions, as mtDNA CN is shifting relative to *MT-ND4*. This suggests these samples may contain deletions (He *et al.*, 2002; Krishnan *et al.*, 2007; Grady *et al.*, 2014a) which will need to be investigated further. An additional factor to consider measuring simultaneously when measuring mtDNA CN would be mitochondrial-related replication protein levels, such as TFAM. More accurate mtDNA CN levels could then be determined relative to replication machinery levels.

My results do not show a gender effect as mtDNA CN changes, which has also been reported between genders in mouse models simulating mtDNA CN replication (Ylikallio *et al.*, 2010). Although mtDNA CN changes between genders has been vastly reported in the literature, continued conflicting reports are being published. This may be a result of tissue-specificity. For example, an age-related decline in mtDNA CN is reported in pancreatic beta cells between genders (Cree *et al.*, 2008a), however blood cells have a much quicker turnover than pancreatic beta cells. The high renewal of blood cells may be masking the real gender effects recorded. Lopez and colleagues have reported a gender-specific genetic regulation of mtDNA CN using a genome wide linkage analysis approach (López *et al.*, 2012), however their work does not include any further functional analysis to support the data.

Other studies in mice have reported gender specific differences in mitochondrial biogenesis, morphology, respiratory function, and ROS homeostasis (Khalifa *et al.*, 2017). Khalifa and colleagues report that oxidative stress biomarkers are lower in healthy young women in age-matched males (Khalifa *et al.*, 2017). This work is further supported by Borrás and colleagues work, where they report that female mice express higher levels of antioxidant genes than male mice (Borrás *et al.*, 2003). Work conducted by Kim and colleagues also show that cyclophilin D is an essential component of oestrogen-mediated neuroprotection for ALS sufferers (Kim *et al.*, 2012), and because ALS is more common in males, this introduces biological protective factors which may be protecting mtDNA from long-term damage.

Malik and colleagues hypothesise that oxidative stress contributes to mtDNA CN changes, as well as mitochondrial number (Malik *et al.*, 2011; Malik and Czajka, 2013). A study in dialysis patients showed a strong correlation between mtDNA CN and oxidative stress (Chen *et al.*, 2008), as was a study conducted by Chan and colleagues (Chan *et al.*, 2013), and other studies in T2DM have reported the same (Beisswenger *et al.*, 2005; Malik *et al.*, 2009), thus suggesting oxidative stress during the immune response might be causing changes in mtDNA CN in cell components. Liu and colleagues also support this theory, as they reported an increase in mtDNA CN in WBC as redox factors such as retinol, non-protein thiols, and ferritin in blood plasma increased. They also reported a decrease in mtDNA CN as vitamin E levels increased (Liu *et al.*, 2003). All these results may support why we reported a statistically significant decrease in mtDNA CN in males with age, but not in females, however further research would need to be conducted to investigate the biological signficiance of these differences.

Measuring ROS or lactate levels directly in cells may act as a useful reference variable when calculating mtDNA CN in tissue. As in the diagnosis of mitochondrial disorders (section 1.2.4.1), where elevated LDH levels in blood serum are an indicator of mitochondrial dysfunction (Robinson, 2006; Haas *et al.*, 2008; Magner *et al.*, 2011), LDH levels are also the strongest prognosis factor for survival rate for patients receiving treatment for metastatic melanoma, and are used in the American Joint Committee on Cancer stage five classification (Balch *et al.*, 2001). Furthermore, LDH, and inflammation levels have been monitored in sepsis patients by Domingos Corrêa and colleagues, whereby they report changes in mitochondrial function which enables certain tissues to recover more readily from sepsis than other tissues

(Corrêa *et al.*, 2017). These clinical applications may also support how mitochondrial stress leads to mtDNA CN differences reported in respective cells, but also should be used to determine likely progress during disease.

Expanding on this, mitochondrial haplogroups have been reported to reduce risk of developing PD, and have a higher sepsis survival rate (Pyle *et al.*, 2005; Lorente *et al.*, 2013). This suggests that mtDNA variants may have more protective or detrimental implications when mtDNA is exposed to biological stresses, and may contribute to mtDNA CN regulation (which is investigated in Chapter 1). Therefore it is possible to hypothesise that mtDNA variants may alter the ability for replication machinery to bind and function under certain conditions more than that of other variants. Which in turn could alter the redox state the cell is performing under, leading to further downstream signalling cascades to alter mitochondrial biogenesis. For example, changes in redox state trigger mitochondrial fusion events, and the efficiency at which mitochondria can accommodate would determine the fate of the mitochondrion and cell (Meyer and Bess, 2012). Indications of this are already reported in neutrophils of people with SLE, whose neutrophils, which are exposed to autoantibodies, release mtDNA instead of undergoing mitophagy (Caielli *et al.*, 2016). This may have long term mtDNA CN regulation defects.

Previous literature has reported differences in blood composition between males and females (Bain *et al.*, 1984; Bain, 1996). UKBS samples have not had blood composition differences reported between genders in their original studies (Gieger *et al.*, 2011; Nalls *et al.*, 2011; Van Der Harst *et al.*, 2012), however my results support the literature, showing significant differences in blood composition between genders in a large European cohort (Figure 3.12 and Figure 3.13). What's more is these results show significant mtDNA CN changes in blood subcomponents, which is dependent on overall blood cellular composition (Figure 3.14 and Table 3.4).

One other study in the literature has investigated mtDNA CN in peripheral blood in a Flemish population of 689 persons. Knez and colleagues work reports a significant age-related decrease in mtDNA CN in their population, and mtDNA CN is significantly higher in females when they adjusted for age. What's more is they report a significant decrease in mtDNA CN as WBC count increased, and this is consistent across WBC subcomponents including neutrophils, monocytes, and lymphocytes. They did not report any significant changes in mtDNA CN in eosinophils or basophils. On the

contrary though, they did report a significant increase in mtDNA CN as platelet count increased (Knez *et al.*, 2016). Their results support the results recorded in this study, because mtDNA CN did significantly decrease with age, but only in males. My results also report significant decreases in mtDNA CN as WBC count increased, but this was strongly associated to neutrophils. No significant decreases in mtDNA CN were associated in monocytes or lymphocytes. In fact, significant increases in mtDNA CN are reported as basophil proportion, lymphocytes, eosinophils, and monocytes increase in my work, and there was no significant change relative to platelet count (Figure 3.14 and Table 3.4).

Despite contrary belief by Dr Rushton, who believes we are only sampling populations of iron deficient women (Rushton *et al.*, 2001), significantly lower haemoglobin and RBC reported in women (Figure 3.12Ai and ii) are also reported in the literature, and is believed to be a result of the menstrual cycle. Changes in blood composition may be an indicator of how well mitochondria are able to adapt under times of high redox, or immune stress. As previously mentioned, elevated neutrophils, monocytes, and LDH levels in blood are associated with a poor survival rate in patients with metastatic renal cell carcinoma (Schmidt *et al.*, 2005). These elevated LDH levels may be an indicator of mitochondrial stress, as cells are trying to neutralise increased lactate levels as a result of inefficient aerobic respiration or mitochondrial trafficking. Increased lactate production is also apparent in blood from patients with mitochondrial diseases, which would support this theory (Robinson, 2006; Haas *et al.*, 2008; Magner *et al.*, 2011).

Finally, I was not able to detect heritability of mtDNA CN between mother and child. This suggests mtDNA CN is not solely genetically controlled and supports the literature that mtDNA CN is biologically regulated, and is not a fixed outcome like that of inheriting your mtDNA genome. This result also encourages the need for genome-wide association studies to identify both nuclear and mitochondrial genes associated with mtDNA CN differences between generations, and investigate biological mediators further.

In conclusion, the results reported in this chapter demonstrate that demographic factors such as age, gender, and blood count composition contribute to different mtDNA CN values observed in DNA samples extracted from blood, and therefore are important covariates that should be included in future studies.

Chapter 4 Nuclear Genetic Factors Modulating Mitochondrial DNA Copy Number

4.1 Introduction

4.1.1 Nuclear - Mitochondrial Intergenomic Communication

As well as being the source of cellular energy production, mitochondria are critical for managing several other functions involved in cellular homeostasis, including the synthesis of nucleotides, amino acids, lipids, and haem, and regulating calcium signalling to control cellular apoptosis, and other signalling events (Schatz, 1995; Duchen, 2000; Wallace, 2005b; Wallace, 2013).

Mitochondrial biogenesis is managed by highly coordinated communication systems between nuclear DNA (nDNA) and mitochondrial DNA (mtDNA) (Figure 4.1), and involves a complex array of signalling networks to maintain both cellular and organelle homeostasis (reviewed by (Quiros *et al.*, 2016)).

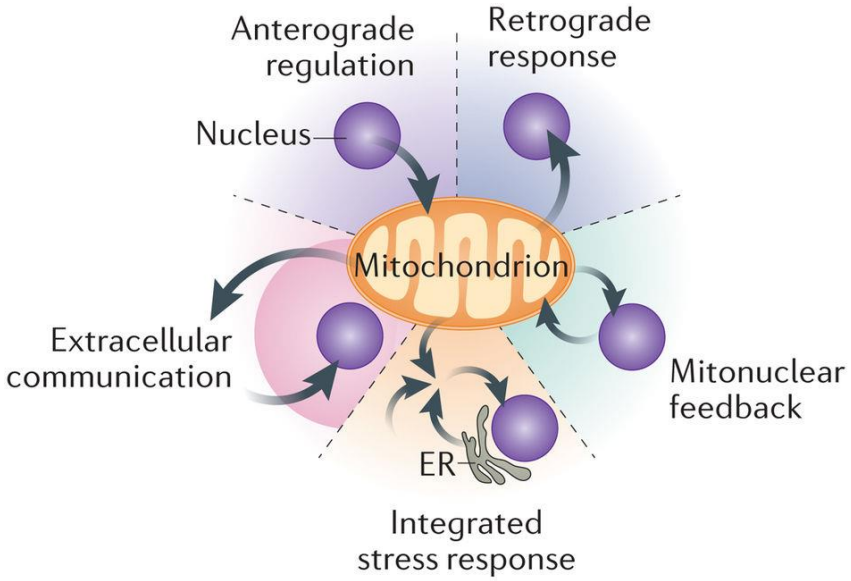


Figure 4.1: An overview of the complex signalling systems between the mitochondria and nucleus. Image replicated from (Quiros et al., 2016).

Mitochondria rely highly on nDNA, as mitochondria require over 1500 nDNA encoded proteins for their maintenance (O'Brien, 2002). Studies in mice which have undergone pro-nuclear transfer have shown that the more divergent nDNA is from the original genetic material, the harder it is for certain tissues to generate sufficient ATP through OXPHOS (McKenzie *et al.*, 2004). Additional studies in mice demonstrated that by altering nuclear-mitochondrial genetic material affected their physical and metabolic performance (Nagao *et al.*, 1998). This is believed to be a result of nDNA struggling to recognise different mtDNA (St. John, 2016).

Genetic divergence has also resulted in sections of mtDNA translocating to nDNA during evolution (Hazkani-Covo *et al.*, 2010; Lang *et al.*, 2012). These sequences are known as nuclear mitochondrial DNA sequences (NUMTs). mtDNA is believed to have migrated in to the nDNA under high stress conditions (Ju *et al.*, 2015; Srinivasainagendra *et al.*, 2017), and to prevent genetics being inherited as a matter of chance through genetic drift (reviewed by (Blanchard and Lynch, 2000)). As a result 15 out of 296 NUMTs are larger than 5.8 kb in size (Mourier *et al.*, 2001), and more than 95% of mtDNA is believed to be duplicated across nDNA (Malik and Czajka, 2013) (Figure 4.2).

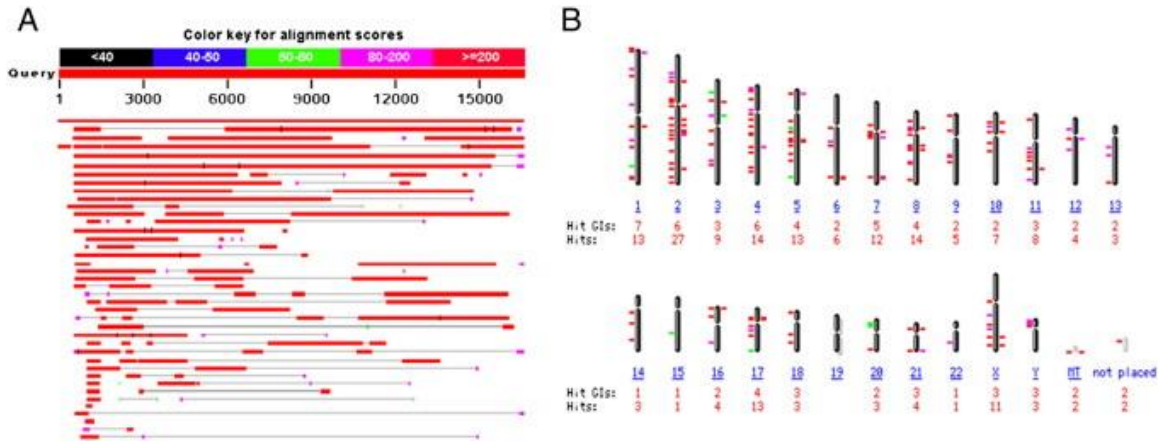


Figure 4.2: [A] Nuclear encoded mitochondrial pseudogenes aligned to corresponding mitochondrial DNA (NC_012920) regions using BLAST. [B] Locations of mitochondrial pseudogenes homology in each nuclear chromosome. Image replicated from (Malik et al., 2011).

Coincidentally, mtDNA CN has also been reported to change when genetic translocations have been reported (Reznik *et al.*, 2016; Srinivasainagendra *et al.*, 2017), however it is unclear whether these changes are a result of disease mechanisms, or genetic translocations.

4.1.2 Nuclear DNA and Mitochondrial DNA Copy Number Regulation

85 of the OXPHOS structural subunits are coded for by nDNA (Anderson *et al.*, 1981; Wallace, 1999; Smeitink *et al.*, 2001), and mammalian tissues have changing bioenergy demands, which results in cells containing different numbers of mtDNA (Robin and Wong, 1988; Capps *et al.*, 2003). There is extensive evidence demonstrating that the relationship between OXPHOS capacity is relative to mtDNA CN (reviewed by (Reinecke *et al.*, 2009)), and that disruptions in OXPHOS machinery cause mtDNA CN changes (Collombet *et al.*, 1997; Bonod-Bidaud *et al.*, 1999; Heddi *et al.*, 1999).

Mitochondrial DNA (mtDNA) represents less than 1% of total genomic DNA present in eukaryotic cells (Clayton, 1982), however mtDNA replication relies entirely on nuclear encoded genes to replicate, and is independent to the cell cycle (Bogenhagen and Clayton, 1977). Mitochondrial transcription factor A (*TFAM*), mitochondrial polymerase gamma (*POLG*), and mitochondrial single-stranded binding protein (*mtSSB*) are nDNA encoded, and are a selection of key proteins involved in mitochondrial replication and transcription, and their expression is regulated by mtDNA CN levels. For example, in mtDNA depleted cells, TFAM levels are slower to recover than mtDNA CN, and this is believed to be a result of mtDNA CN stimulating transcription of *TFAM* (Seidel-Rogol and Shadel, 2002), however this may be regulated by *NRF-1* and *NRF-2*, which are also involved in mitochondrial biogenesis regulation (Virbasius and Scarpulla, 1994). The relationship between TFAM and mtDNA has been extensively researched, and it is believed that TFAM is a molecular chaperone for mtDNA, protecting it against oxidative damage during transcription (Chinnery and Hudson, 2013).

mtSSB is a nDNA-encoded protein which stabilises single-stranded regions of mtDNA at replication forks during mtDNA replication and transcription, and enhances POLG activity (Takamatsu *et al.*, 2002). Loss of mtSSB causes decreased mtDNA replication, and decreased mtDNA CN (Ruhanen *et al.*, 2010).

POLG is the only mitochondrial replication polymerase, and is the main cause for mitochondrial instability disorders (Chinnery and Hudson, 2013). *POLG* has been identified in families with autosomal dominant, or autosomal recessive *PEO1* (*Twinkle* gene) defects, which results in high numbers of mtDNA deletions (Horvath *et al.*, 2006). Additionally, mutations in *POLG* cause diseases such as Alpers Syndrome, whereby reduced mtDNA CN results in progressive cerebral atrophy (Stewart *et al.*, 2011; Copeland, 2012).

Genetic defects in nDNA-encoded genes involved in nucleotide metabolism also cause mitochondrial-related diseases. Mitochondrial depletion syndromes (MDS) are a rare cause of mitochondrial disease, and are characterised by a 30% lower mtDNA CN than normal cells (Rötig and Poulton, 2009), which results in multi-organ failure as tissues rely highly on aerobic respiration (El-Hattab and Scaglia, 2013). MDS are often a result of mutations in nDNA genes such as mitochondrial thymidine kinase 2 (*TK2*), deoxyguanosine kinase (*DGUOK*), and *Twinkle* (*C10orf2*). Mutations in *TK2*,

and *DGUOK* disrupt nucleotide pooling, which reduces mtDNA replication efficiency, therefore severely depleting mtDNA CN (Copeland, 2012; Chinnery and Hudson, 2013). Additionally, mutations in adenine nucleotide translocase 1 (*ANT1*) interrupts mitochondrial membrane ADP-ATP exchange which causes further nucleotide imbalances, and mtDNA CN depletion (El-Hattab and Scaglia, 2013; Nikkanen *et al.*, 2016). MDS has been extensively reviewed in the literature (Rötig and Poulton, 2009; Copeland, 2012).

nDNA and mtDNA inter-genomic communication is extremely sensitive, and Dean and colleagues even report complete loss of mtDNA in African trypanosome *Trypanosoma brucei* parasites as a result of small amino acid changes in nDNA (Dean *et al.*, 2013). Kelly and colleagues also identified the importance of efficient, and co-ordinated nDNA-mtDNA communication during differentiation in mouse embryonic stem cells, whereby nuclear reprogramming lead to unregulated mtDNA CN levels and ATP content, which prevented normal differentiation (Kelly *et al.*, 2013b).

Despite extensive literature reporting genes associated to disease, or mechanisms by which mtDNA CN changes, no study has attempted to identify genes that regulate mtDNA CN in large populations of healthy individuals. Therefore, to address this, I performed a genome-wide association study (GWAS) in a large population of unrelated, healthy, European individuals to identify nuclear genes which are associated with mtDNA CN changes.

4.1.3 A Brief History of Genome-wide Association Studies

A GWAS is an approach that involves finding associations between a particular trait and (usually millions of) genetic markers across DNA (National Human Genome Research Institute, 2003). The large number of genetic markers being tested makes GWAS beneficial for finding genetic variations associated to complex diseases such as type 2 diabetes (Morris *et al.*, 2012; Replication *et al.*, 2014; Fuchsberger *et al.*, 2016), obesity (Locke *et al.*, 2015; Shungin *et al.*, 2015), immune disorders (Petukhova *et al.*, 2010; Okada *et al.*, 2014; Ahola-Olli *et al.*, 2017), and neurological illness (International Multiple Sclerosis Genetics *et al.*, 2011; Lambert *et al.*, 2013; Nalls *et al.*, 2014; consortium, 2015), as well as common traits like height (Wood *et al.*, 2014; Marouli *et al.*, 2017). GWAS have increasingly become more popular in recent years as genetic sequencing has become cheaper, and data are more

accessible. As a result, a catalogue of published GWAS is available (Hindorff LA. *et al.*, 2017).

The introduction of the International HapMap Project in 2003 was also a major contribution to the increase of GWAS. As an international consortium, the HapMap Project was set up to map common genomic patterns of human DNA sequence variation between different populations, in the hope that they could be used to identify sequence variants associated with complex diseases. Genomic patterns were identified by using more than one million sequence variants, their frequencies, and degree of association between them, and the data was freely available to the public (International HapMap, 2003). Since then, the HapMap project has continued to grow, and in 2007 a haplotype reference panel of 420 haplotypes across 3.1 million SNPs in three continental populations was available, and more recently the 1000 Genomes Project has been set up to cumulate a reference panel of 5008 haplotypes over 88 million variants from 26 worldwide populations using datasets from low-coverage whole-genome sequencing (WGS) (Genomes Project *et al.*, 2015). This resulted in projects collating data together to create haplotype reference panels (Genome of the Netherlands, 2014; Sidore *et al.*, 2015), which have been a vital resource to be able to increase genetic power in GWAS, as genotyping data can be imputed using these reference panels (Marchini *et al.*, 2007; Howie *et al.*, 2009; Huang *et al.*, 2015).

GWAS in centenarians across different populations of people have reported inconsistencies in results when searching for loci associated to complex traits, because both environmental and genetic factors contribute to the trait developing (Aulchenko *et al.*, 2009; Garagnani *et al.*, 2014; Wood *et al.*, 2014). For example, the mitochondrial proteome shows tissue specific profiles and has different bioenergetic behaviours (Malik and Czajka, 2013). Several GWAS have identified nuclear genes, involved in mitochondrial biogenesis, to be associated with age-related diseases; such as *PPAR*-γ, *PGC-1α*, and *UCP*s (Jia *et al.*, 2010; Manning *et al.*, 2012; Morris *et al.*, 2012; Wallace, 2013; Garagnani *et al.*, 2014). Garagnani and colleagues discuss how a number of genes involved in the regulation of mitochondrial bioenergetics also contribute to longevity, by looking at nDNA, mtDNA and microbiome DNA interactions (Garagnani *et al.*, 2014).

Several models for testing continuous traits in an association-like study have been suggested (Shin and Lee, 2015; Wu and Pankow, 2015), whereby mixed-models try to accommodate to environmental and genetic variability. However, larger sample sizes are usually needed, which usually require meta-analysis of datasets, but this also has limitations and bias which need to be adjusted for (Egger *et al.*, 1997; Rucker *et al.*, 2011; Mueller *et al.*, 2013). Genome-wide significant threshold is usually set at 5.00E-08 to avoid false positive discoveries due to hundreds of tests being performed across millions of SNPs (Visscher *et al.*, 2014).

4.1.4 Justification for Study

A PubMed search for [mitochondrial copy number] reports more than 2000 publications, however there are no publications whereby a GWAS has been performed to identify genetic variants contributing to mtDNA CN changes. Curran and colleagues used linkage analysis in 42 extended Mexican-American families to try and address this, however they only measured 439 highly polymorphic markers (Curran *et al.*, 2007). Cai and colleagues also performed a GWAS in over 10,000 Chinese individuals, and calculated mtDNA CN as a proxy from low coverage sequencing data (Cai *et al.*, 2015). Their study identified a variant in *TFAM* (p= 8.73E−28), and variant in *CDK6* (p= 6.03E-16) to be associated with mtDNA CN changes in individuals with major depressive disorder (MDD). However, another GWAS conducted in over 5000 Chinese with MDD, could not replicate these results (consortium, 2015). Furthermore, Cai and colleagues did not conduct any real-time PCR assay measurements to validate the accuracy of proxy measurements to actual mtDNA CN measurements, thereby limiting the reliability of the results.

Based on my results in section 3.3.6, the heritability of mtDNA CN is also very low, however other studies have reported heritability to be about 60% (h^2=0.60) (Curran *et al.*, 2007; Xing *et al.*, 2008), therefore it would be expected that a few nuclear genes would be regulating mtDNA CN, and not a single locus as reported by Cai's group (Cai *et al.*, 2015). Therefore to address this, I performed a GWAS in over 8000 healthy, European individuals, using actual measures of mtDNA CN.

4.2 Aim

To identify nuclear genetic loci associated with mtDNA CN changes by performing a GWAS in a large, healthy, unrelated, European population. Then a meta-analysis will be performed in an independent, healthy, unrelated, European population to validate, and replicate, the loci identified in the discovery study.

4.3 Materials and Methods

4.3.1 Cohorts

4.3.1.1 Cohort One - UKBS

Full cohort details and methods are available in Chapter 2. For work in this chapter, 3091 UKBS DNA samples which had been extracted from white blood cell pellets from blood using phenol-chloroform methods were provided to Newcastle University, and FBC data was provided for 2617 samples through collaboration with Professor Willem H. Ouwehand at University of Cambridge. Relative mtDNA CN was determined using an established triplex TaqMan™ qPCR assay, quantifying mitochondrial genes *MT-ND1* and *MT-ND4*, to nuclear housekeeping gene *B2M* (Grady *et al.*, 2014a; Pyle *et al.*, 2015b; Rygiel *et al.*, 2015). 2735 samples had 1,152,686 variants successfully genotyped on the on the Illumina 1.2M Duo array (Illumina, San Diego, California, USA), of which 2682 (n=1336 males, n=1346 females) had both genotyping and mtDNA CN data.

4.3.1.2 Cohort Three - ALSPAC

Full cohort details and methods are available in Chapter 2. For work in this chapter, 5461 ALSPAC mothers were used as a replication cohort to this study. DNA for qPCR analysis was either extracted from whole blood (37.6%) or WBC (62.4%) using a phenol-chloroform method (Fraser *et al.*, 2013), and estimated blood cell count data was generated for 546 samples. Relative mtDNA CN was determined using a singleplex SYBR® Green assay which quantified a region of the mitochondrial D-loop to nuclear housekeeping gene *B2M.* All samples were additionally genotyped on the Illumina Human660W-Quad array (Illumina, San Diego, California, USA), and all samples had both SNP genotyping and mtDNA CN data. All analysis for cohort three was conducted at University of Bristol, and summary data was provided to Newcastle University to perform respective meta-analysis'.

4.3.2 Pre-Imputation

4.3.2.1 Cohort One - UKBS

Following advised quality-control check protocols (Anderson *et al.*, 2010), and in line with previous GWAS (Burton *et al.*, 2007; Hudson *et al.*, 2013b; Fuchsberger *et al.*, 2016), genotyping data was cleaned, and quality control checks were performed using PLINK v1.90b3p 64-bit (2 Aug 2014) (Purcell *et al.*, 2007).

4.3.2.1.1 Genotype Filtering

Before QC, samples had an average missingness rate of 0.0023 (±0.0058 s.d), and 11 individuals were removed for missingness >5%. 4,249 variants were removed due to missing genotype call rates >5%, 12,344 variants were removed due to Hardy-Weinberg threshold 1.00E-06, and 187,314 variants were removed due to minor allele threshold <1% (Burton *et al.*, 2007). 2724 individuals (1358 males, 1366 females) and 948,779 variants passed QC, with a total genotyping rate of 0.997998, and average missingness data rate of 0.0016 (±0.0048 s.d). Outliers for autosomal heterozygosity were calculated as 99.60%. No duplicate SNPs were detected as they were removed in the original study (Burton *et al.*, 2007). Genotyping data was then restricted to autosomes, and pseudo-autosomal regions of X were removed. This removed 27,323 variants, with 921,456 variants remaining.

4.3.2.1.2 Identical-by-Descent Testing

Clustering analysis was performed and one single cluster was identified. Identical-by-descent (IBD) analysis was performed across default size, 500kbp regions, at the standard threshold 0.85, then 0.2, and then without a threshold. No individuals were IBD at any threshold, which is expected as samples were cleaned in the original study (Burton *et al.*, 2007).

4.3.2.1.3 Principal Component Analysis

Multi-dimensional scaling was originally used to clean UKBS genotyping data in the original study (Burton *et al.*, 2007), however further principal component analysis was conducted to detect for any individuals with non-European ancestry. PCA analysis was performed using the '—PCA' command in PLINK, and the top 20 principal components of the variance-standardized relationship matrix were tested, generating respective eigenvalues and eigenvectors (Figure 4.3).

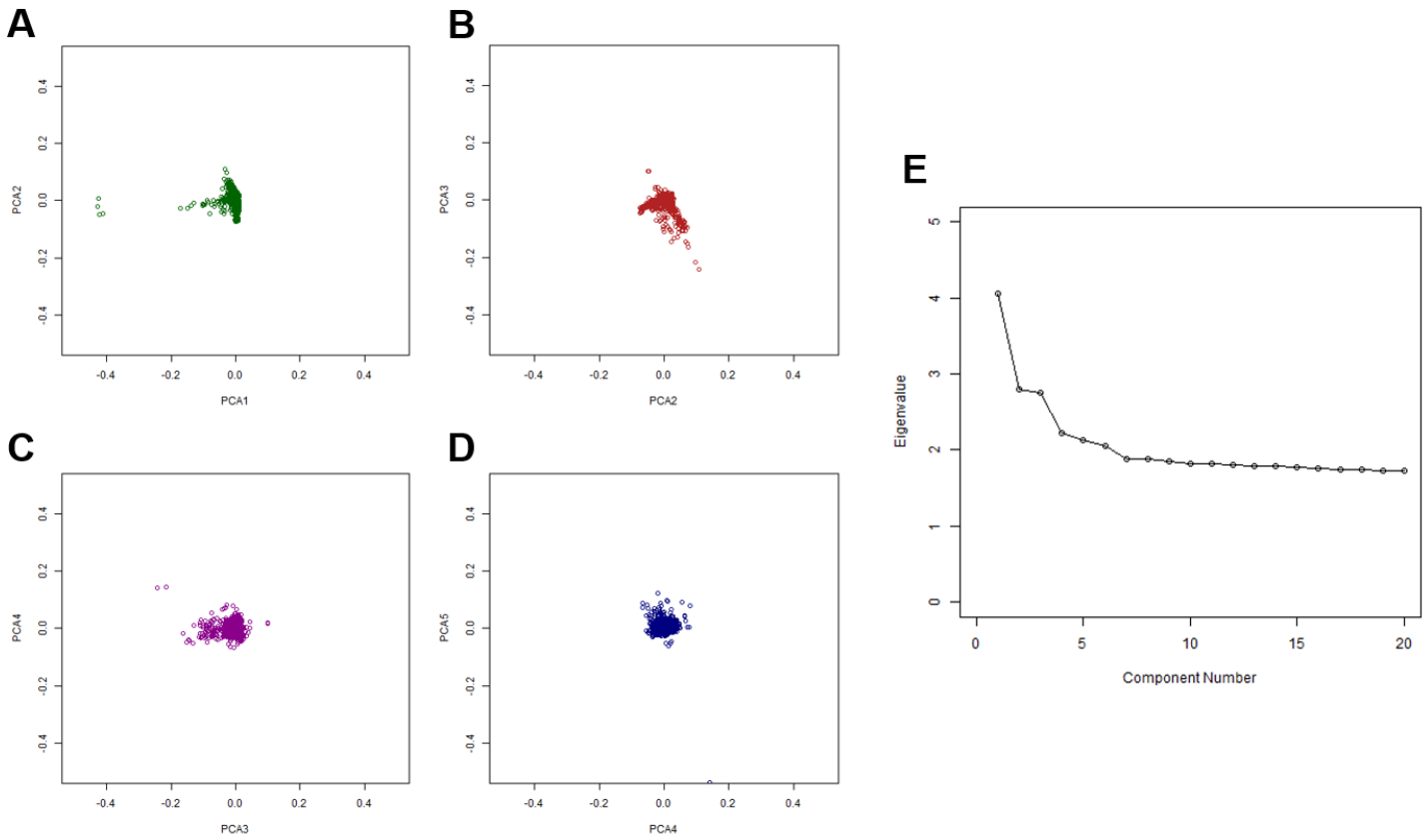


Figure 4.3: Multidimensional scaling analysis on UKBS samples (n=2724). Cluster plots of eigenvectors [A] one and two, [B] two and three, [C] three and four, and [D] four and five. [E] Scree plot of each eigenvalue for each respective principal component number.

4.3.2.1.4 Preparing Files to be Imputed

To increase the genetic power of performing a GWAS on UKBS samples, SNP genotyping were imputed. Two imputation servers were available; the Sanger Imputation Server (McCarthy *et al.*, 2015), and the Michigan Imputation Server (Das *et al.*, 2016). As I had not performed imputation before, and no literature was published comparing imputation servers, I decided to impute UKBS genotyping data using both servers, and then compare results to identify which would be most suitable to perform my GWAS analysis.

Both servers required genotyping alleles to be on the forward strand, and in the GRCh37-hg19 genome build format. UKBS genotyping data was sequenced to the NCBI36-hg18 genome build therefore this needed to be updated. Both servers advised using Will Rayner's toolbox (http://www.well.ox.ac.uk/~wrayner/tools/) to perform strand alignments, and to 'LiftOver' genotyping data to the correct genome build before HRC imputation.

4.3.2.1.4.1 Strand Alignment

Strand alignments assume that genotyping calling algorithms use the Illumina top strand alleles. Despite UKBS samples being sequenced on an Illumina genotyping array system, strand alignment was performed to confirm accuracy.

Recommended strand alignment instructions (http://www.well.ox.ac.uk/~wrayner/strand/) were used, and the 'update_build.sh' was downloaded along with the NCBI36-hg18 reference strand data - Human1.2MDuoCustom_v1_A-b36-strand.zip. Reference data was unzipped, and the final few rows of the strand file were removed, as they caused complications later in the data conversion. The command to action the strand alignment was: <update_build.sh> <UKBS_cleaned_genotype_file> <Human1-2M-DuoCustom_v1_A-b36.strand> <output_file_name>. This generated three new binary files which had been updated to the forward strand alignment.

4.3.2.1.4.2 LiftOver

Both imputation servers use the more recent GRCh37-hg19 build of the human genome. Therefore the strand aligned data files were converted ('lifted') from the NCBI36-hg18 build to the GRCh37-hg19 build using UCSC's Lift Genome Annotation (LiftOver) online conversion tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). Bim files can be converted using the online platform, or by downloading reference

sequences, and then manually update bim files in Unix. Both methods produced the same results, however I used the online tool for this study.

Before performing any conversions, strand aligned, binary bim files were converted to an alternative 'bed' format that LiftOver could process (Figure 4.4). The new 'bed' format had to contain: chromosome number (as chrN), start base position and end base position of the variant (both written as a single number), and SNP ID. I also included major and minor alleles in the new 'bed' format. Once genotyping was in the 'bed' format, genotyping lift over was performed.

Bim Format

CHR	SNP_ID	CTMN	BP	A1	A2
1	rs9645429	0	569604	G	A
1	rs3094315	0	752566	G	A
1	rs3131972	0	752721	A	G
1	rs3131969	0	754182	A	G
1	rs1048488	0	760912	C	T
1	rs12562034	0	768448	G	A
1	rs4040617	0	779322	A	G
1	rs2980300	0	785989	T	C
1	rs11240777	0	798959	G	A
1	rs4970383	0	838555	C	A

'Bed' Format

CHR	START_BP	END_BP	SNP_ID	A1	A2
chr1	569603	569604	rs9645429	G	A
chr1	752565	752566	rs3094315	G	A
chr1	752720	752721	rs3131972	A	G
chr1	754181	754182	rs3131969	A	G
chr1	760911	760912	rs1048488	C	T
chr1	768447	768448	rs12562034	G	A
chr1	779321	779322	rs4040617	A	G
chr1	785988	785989	rs2980300	T	C
chr1	798958	798959	rs11240777	G	A
chr1	838554	838555	rs4970383	C	A

Figure 4.4: Example of converting a bim file (top) to a 'bed' file (bottom) format for LiftOver.

Using the online platform; human and Mar. 2006 (NCBI36/hg18) was selected as the original human genome assembly build, and human and Feb. 2009 (GRCh37/hg19) was selected as the new genome assembly build. The minimum ratio of bases that must remap was set to the default of 0.95. The new 'bed' format genotype file was then uploaded to the online platform and submitted. A new genome build file was then generated which was downloaded. 897,705 variants (97.42%) were successfully updated to the GRCh37-hg19 genome build.

Following conversion, the UKBS converted GRCh37-hg19 build genome file had '#' notes removed, and was manually converted back in to the correct bim format (Figure 4.4). The original bim file was then filtered against the new bim file to generate a list of variants which were not converted. This list of non-converted variants was then used to filter the original, strand-aligned, binary files in PLINK, producing binary files which contained 897,705 variants which had been strand-aligned, and converted to the GRCh37-hg19 genome build. These variants were then sent for imputation.

4.3.2.2 Cohort Three - ALSPAC

19,930 individuals (mothers and children) entered the quality control pipeline, including 526,688 variants for mothers. Samples were removed with an average missingness rate >5%, and if they had indeterminate X chromosome heterozygosity, or extreme autosomal heterozygosity. Variants were removed with missing genotype call rates >5%, MAF <1%, and a Hardy-Weinberg equilibrium p value <5.00E-07. Multidimensional scaling of genome-wide identical by state pairwise distances, using the four HapMap populations as a reference, removed samples showing evidence of population stratification. Cryptic relatedness was assessed, and 5461 mothers were unrelated with an IBD >0.125. 465,740 variants passed QC, however a further 112 were removed when genotyping was converted to the GRCh37-hg19 build, and another 234 variants were removed due to Hardy-Weinberg Equilibrium. 465,394 variants were finally used for imputation.

4.3.3 Imputation

To increase genetic power, both UKBS and ALSPAC cohorts were imputed using the Haplotype Reference Consortium (HRC) reference panel. HRC release 1.0 was formed in 2015, and consisted of 64,976 haplotypes of predominantly European ancestry from 39,235,157 SNPs (McCarthy *et al.*, 2015). Release 1.1 now contains 64,940 haplotypes, generated from 40,176,563 SNPs and 32,470 samples (McCarthy *et al.*, 2016). Imputing European cohorts is genetically advantageous because European-ancestry populations have longer haplotype blocks containing highly correlated SNPs, therefore fewer genotyped SNPs are needed for genome-wide coverage (Edwards *et al.*, 2013).

ALSPAC samples were imputed to the HRC release 1.0 reference panel before our collaboration was organised. Therefore, to maintain consistency between analysis,

UKBS samples were imputed to the HRC release 1.0 (r1.0), and HRC release 1.1. (r.1.1) reference panels. Imputation results could then be compared between HRC r1.0 and r1.1 to identify quality differences.

4.3.3.1 Cohort One - UKBS

Following QC, liftover, and strand alignment measures, 897,705 genotyped variants were sent for imputation.

4.3.3.1.1 Sanger Imputation Server

The Sanger Imputation Service is provided by the Wellcome Trust Sanger Institute (Hinxton, UK), and was introduced as part of the first release of the HRC (McCarthy *et al.*, 2015). UKBS cleaned genotyping binary files were converted in to VCF format using PLINK v1.90b3p 64-bit (2 Aug 2014) (Purcell *et al.*, 2007) and submitted to the server. Genotyping data was phased using EAGLE2, which estimates haplotypes from the reference panel (Browning and Browning, 2011; Durbin, 2014; Loh *et al.*, 2016), and imputation was conducted against the HRC reference panel using Positional Burrows-Wheeler Transform (PBWT) methods (https://github.com/VertebrateResequencing/pbwt), which store and compute on genome variation data sets. UKBS genotyping data was submitted twice to the Sanger Imputation Server to allow genotyping to be imputed to HRC r1.0, and HRC r1.1 reference panels. Following imputation, VCF files were downloaded from the server using Globus data sharing.

4.3.3.1.2 Michigan Imputation Server

Michigan Imputation Server is a remote server which is powered by Cloudgene and supported by the U.S. National Institutes of Health and the FWF Austrian Science Fund. UKBS cleaned genotyping binary files were separated in to separate chromosome files (1-22), and then were individually converted to VCF format using PLINK v1.90b3p 64-bit (2 Aug 2014) (Purcell *et al.*, 2007). All VCF files were submitted to the imputation server, and genotyping data was phased using EAGLE2 (Browning and Browning, 2011; Durbin, 2014; Loh *et al.*, 2016). Imputation was conducted against the HRC reference panel using Minimac3 methods, which use low computational memory to implement imputation algorithms (Das *et al.*, 2016). UKBS genotyping data was submitted twice to the Michigan Imputation Server to allow genotyping to be imputed to HRC r1.0, and HRC r1.1 reference panels. Following imputation, VCF files were downloaded directly from the server.

4.3.3.1.3 Sanger versus Michigan Imputation

Both imputation servers imputed 39,235,157 variants from the HRC r1.0 reference panel, and 39,131,578 variants from the HRC r1.1 reference panel. To compare the quality of each imputation server method, MAFs and imputation scores were compared.

4.3.3.1.3.1 Minor Allele Frequencies

MAFs were generated from VCF files containing genotyped and imputed variants using PLINK 1.9b3.44. Scatter plots of MAFs for each variant were plotted using R studio version 3.0.1 (2013-05-16), whereby MAFs from Sanger-imputed data were compared to Michigan-imputed data (Figure 4.5).

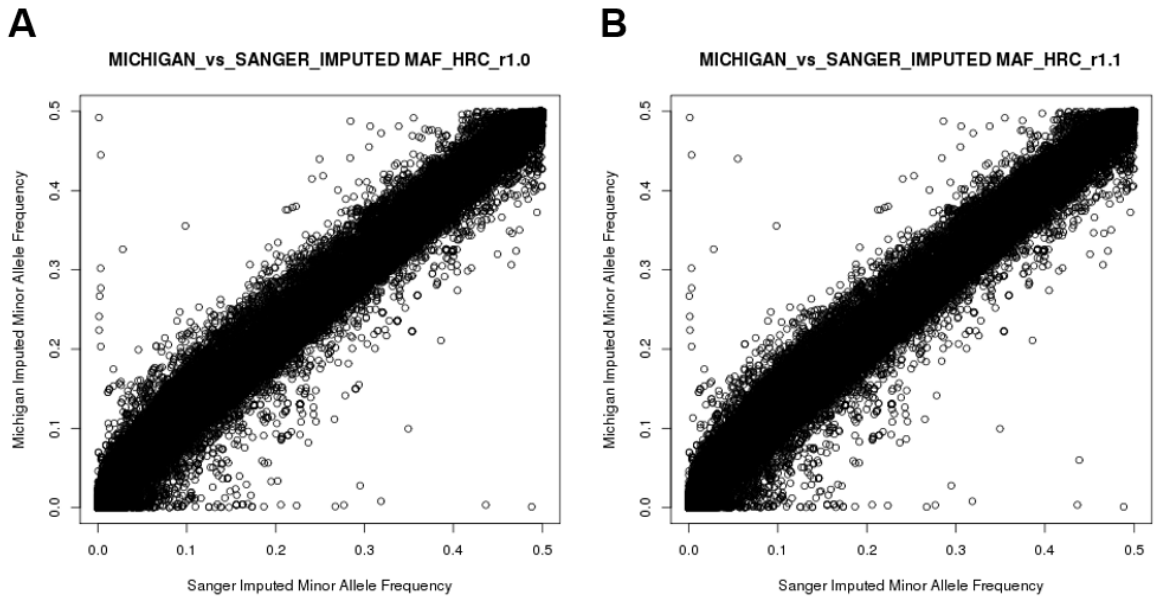


Figure 4.5: Scatter plot comparing the minor allele frequencies of variants from imputed UKBS genotyping data using the Sanger Imputation Server and Michigan Imputation Server. [A] Imputed data using HRC release 1.0 reference panel. [B] Imputed data using HRC release 1.1 reference panel.

As Figure 4.5 shows, MAFs were positively correlated between imputation servers, and there was no difference in variant MAFs when using the HRC r1.0 (Figure 4.5A) or HRC r1.1 (Figure 4.5B) reference panels, which suggested the variants in the reference panels were very similar.

4.3.3.1.3.2 Imputation Scores

Sanger and Michigan Imputation Servers used different methods to impute genotyping data. Sanger used PBWT whereas Michigan used Minimac3. To identify the quality of imputation between Michigan and Sanger Imputation Servers, imputation scores were extracted from imputed VCF files using vcftools version 0.1.12b, and using chromosomes 21 and 22 as an example, imputation scores were compared (Figure 4.6).

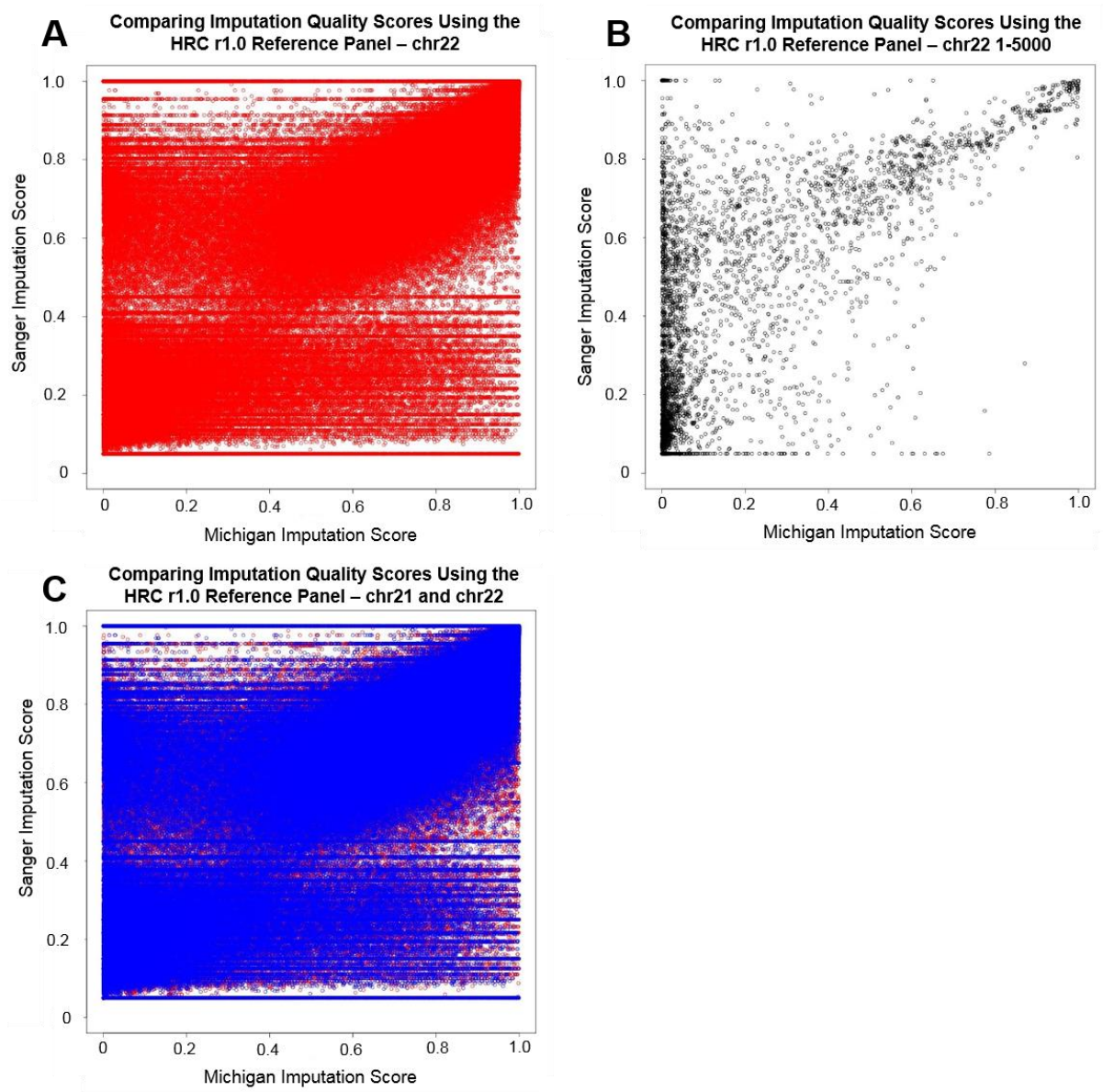


Figure 4.6: Scatter plots comparing imputation scores for each respective SNP that was imputed from the HRC release 1.0 reference panel using either the Michigan Imputation Server (x-axis) or Sanger Imputation Server (y-axis). [A] All chromosome 22 UKBS imputed SNPs. [B] First 5000 imputed SNPs for chromosome 22. [C] All chromosome 21 (blue) and 22 (red) UKBS imputed SNPs overlaid.

As Figure 4.6 shows, imputation scores between Sanger and Michigan Imputation servers were not correlated, and were very different. The imputation score differences seemed to be a result of imputation method differences. This is because the same imputation score pattern replicated between chromosomes 22 (Figure 4.6A), and 21 (Figure 4.6C). All chromosomes are not displayed because images are too noisy and unclear to interpret. Figure 4.6B gives a more detailed view of how imputation scores were so different, as only variants 1-5000 on chromosome 22 were graphed. The same results were reported for HRC r1.1 imputed variants.

4.3.3.1.3.3 Variants Remaining Using Different Filters

To investigate the effect different imputation score and MAF filters had on the number of imputed variants that passed QC, both Sanger and Michigan imputed datasets were filtered under different conditions (Figure 4.7). Imputed data was filtered to include variants with a certain imputation score (Figure 4.7A). All imputed data was then filtered to remove variants with a MAF <0.01, and then were filtered again using different imputation score conditions (Figure 4.7B).

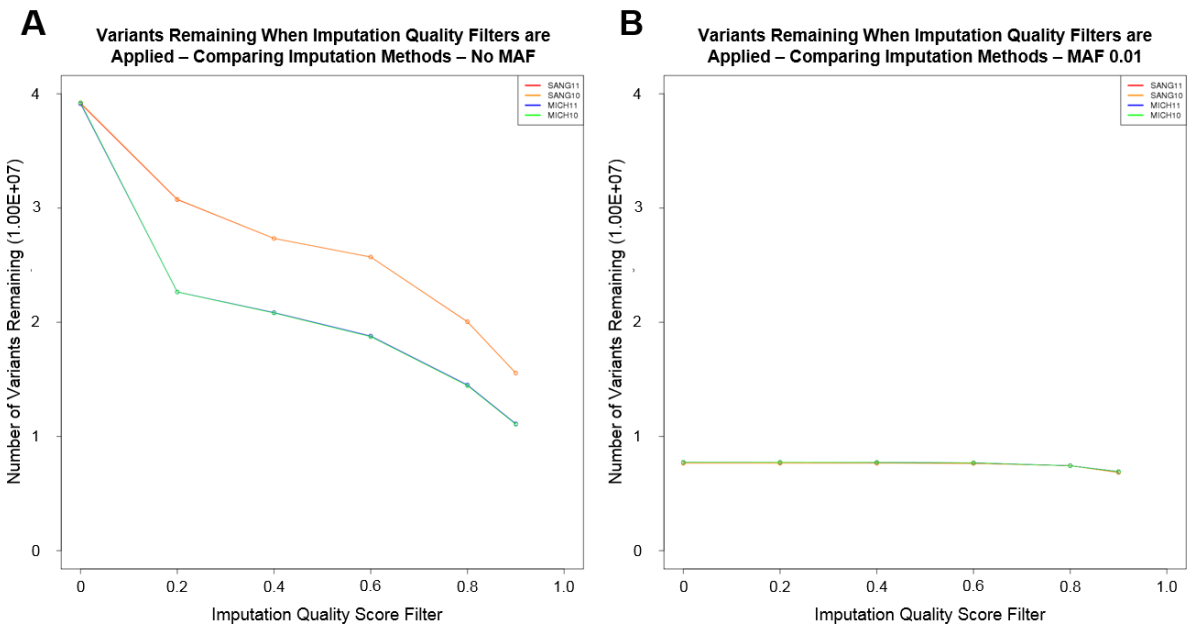


Figure 4.7: Line graphs depicting the number of variants remaining (y-axis) when different imputation quality scores (x-axis) were applied as filters to UKBS genotyped data that had been imputed using the Sanger Imputation Server or the Michigan Imputation Server, referenced to HRC r1.0 or HRC r1.1 reference panels. [A] No MAF filter included. [B] Variants also filtered for MAF< 0.01.

As Figure 4.7A shows, Michigan-imputed variants had more variants removed under certain imputation score filter conditions compared to Sanger-imputed variants when no MAF filter was applied. When a MAF <0.01 filter was applied, Sanger and Michigan-imputed variants had similar numbers of variants removed under different imputation score filters (Figure 4.7B). These data suggest that the Sanger Imputation Server is more effective at imputing rare variants (MAF< 0.01) than the Michigan Imputation Server because more variants are included under different imputation score filters in Sanger-imputed data when no MAF filter is applied, but this is removed when filtering MAF< 0.01 is applied. No differences were reported between HRC r1.1 and HRC r1.0 imputed datasets.

To investigate this in more detail, chromosome 22 was isolated from all imputed datasets (because it is the smallest chromosome to analyse), and imputation score and MAF filters were applied, and then the number of variants remaining was recorded (Table 4.1).

Filters	Sanger_r1.1	Sanger_r1.0	Michigan_r1.1	Michigan_r1.0
No_filters	524,544	524,731	524,544	524,731
INFO>0.2	410,299	410,269	304,721	304,735
INFO>0.4	360,691	360,612	278,812	278,817
INFO>0.6	336,416	336,313	248,140	248,137
INFO>0.8	256,975	256,888	187,007	186,980
INFO>0.9	198,730	198,609	142,116	142,091
MAF>0.01	105,299	105,291	106,756	106,756
MAF>0.05	74,665	74,655	75,005	75,005
MAF>0.10	58,439	58,436	58,603	58,604
MAF>0.01 INFO>0.6	104,944	104,934	105,669	105,669
MAF>0.01 INFO>0.8	101,141	101,143	100,824	100,822
MAF>0.01 INFO>0.9	92,087	92,095	92,880	92,882
MAF>0.05 INFO>0.6	74,501	74,490	74,674	74,674
MAF>0.05 INFO>0.8	73,498	73,497	73,482	73,483
MAF>0.05 INFO>0.9	71,031	71,028	71,108	71,108

Table 4.1: Number of chromosome 22 imputed variants remaining, when different quality control filters are applied. Sanger and Michigan imputation server methods are compared when using reference panel HRC release 1.0 or HRC release 1.1. INFO= imputation score. MAF= minor allele frequency.

Table 4.1 replicates the results reported in Figure 4.7, as the number of variants remaining when imputation score filters alone were applied was very different between imputation servers. However, when MAF filters were applied, the number of variants remaining were very similar between servers. The number of variants remaining between imputation servers was more similar the more stringent the MAF filter applied. Results for HRC r1.0 and HRC r1.1 imputed datasets were similar between imputation server methods.

Subsequent to this analysis, and following previously established methods to perform GWAS using imputed genotyping data, UKBS imputed data was filtered to remove variants with MAF <0.01 and imputation score <0.80 (Marchini *et al.*, 2007). The '—keep-allele-order' command was included in all filtering steps to ensure correct reference allele assignment was maintained. 7,441,490 variants, and 2724 individuals (1358 males, 1366 females) passed QC (Table 4.2).

4.3.3.2 Cohort Three - ALSPAC

All genotyping and imputation was performed by Dr Gibran Hemani (University of Bristol), Dr George McMahon (University of Bristol), and Professor Jonathan Marchini (University of Oxford). Genotyped data was phased using ShapeIT (version 2.r644) (Delaneau *et al.*, 2012; Delaneau *et al.*, 2014; O'Connell *et al.*, 2014), and then imputed to HRC version 1.0 reference panel (McCarthy *et al.*, 2016) using IMPUTE (version 3) (Marchini *et al.*, 2007; Howie *et al.*, 2009). Following imputation, variants with MAF <0.01 and imputation score <0.80 were removed. 7,401,334 variants and 5461 mothers passed QC (Table 4.2).

Stage	Metric		Cohort	
			ALSPAC Mothers	***UKBS***
Pre-imputation	Genotyping platform		Illumina Human660W-Quad	Illumina 1.2M Duo
	Software for QC		PLINK v1.07	PLINK v1.90b3p
	SNP-wise filters	Call rate	0.05	0.05
		Hardy Weinberg Equilibrium	p < 5.00E-07	p < 1.00E-06
		Minor Allele Frequency	0.01	0.01
	Sample-wise filters	Individual call rate	0.05	0.05
		Outliers for autosomal heterozygosity	Removed	Removed
		Samples with indeterminate X chromosome heterozygosity	Removed	Removed
		Cryptic relatedness	IBD >0.125	IBD >0
		Population	Samples removed if clustered outside of CEU HapMap2, using multidimensional scaling	See Burton et al., 2007
Imputation	Haplotype estimation and phasing		ShapeIT (v2.r644)	EAGLE2 (v2.0.5)
	Imputation software		IMPUTE (v3)	PBWT (Sanger Imputation Server)
	Imputation panel		Haplotype Reference Consortium (HRC) Reference Panel version 1.0	
Post-imputation	SNP-wise filters	Minor Allele Frequency	0.01	
		Imputation quality	0.8	
	Number of variants included		**7,401,334**	**7,441,490**
	Number of samples included		**5461**	**2671**

Table 4.2: Summary of methods used at Newcastle University and University of Bristol to prepare UKBS and ALSPAC genotyping data for meta-analysis.

4.3.4 Genome-wide Association Analysis

To ensure consistency in comparing analysis between UKBS and ALSPAC cohorts, genotyping data imputed to HRC r1.0 was used to match ALSPAC data. Data imputed using the Sanger Imputation Server was used for all GWAS analysis because more variants remained following QC filters. Linear additive models were tested using PLINK v1.90b3p 64-bit (2 Aug 2014) (Purcell *et al.*, 2007), using expected allelic score dosage data for ALSPAC, and best guess format data for UKBS. The threshold for genome-wide significance was set at p= 5.00E-08.

4.3.4.1 Cohort One - UKBS

4.3.4.1.1 Pre-imputation

Initial GWAS analysis was performed on NCBI36-hg18 genome build SNP genotyped data which had autosomes removed, and had been cleaned using filters detailed in section 4.3.2.1. 921,456 variants and 2724 individuals (1358 males, 1366 females) were included in linear association analysis using logged, z-scored mtDNA CN as a continuous phenotype. 2671 individuals had phenotype and genotype data.

4.3.4.1.1.1 Covariates

Basic analysis included four PCAs and panel as covariates to adjust for population heterogeneity, and genotyping batch effects. Gender, age, and blood count information including; absolute eosinophil count (EA), absolute monocyte count (MA), absolute lymphocyte count (LA), absolute neutrophil count (NA), and platelet counts (PLT) were included as extra covariates to investigate their effect on association test results. Four PCAs were determined as most appropriate to include in all analysis (Figure 4.3). Blood count information was available for 2305 individuals, however 2292 individuals had additional platelet count data.

4.3.4.1.2 Post-imputation

GWAS analysis was conducted on genotyping data which was imputed using the Sanger Imputation Server on the GRCh37-hg19 genome build to the HRC version 1.0 reference panel (McCarthy *et al.*, 2016). Imputed variants with a MAF <0.01 and imputation score <0.80 were removed, and then linear additive association tests were performed using logged z-scored mtDNA CN as a continuous phenotype. A further filter of MAF <0.01 was included in all association testing, and 95% confidence intervals were included in output files using the '—ci' command. 7,441,490 variants

and 2724 individuals (1358 males, 1366 females) passed QC and were included in analysis when analysing all UKBS data. 2671 individuals had phenotype data. As a discovery study, genome-wide significant (GWS) threshold was set at *p* <5.00E-08.

4.3.4.1.2.1 Stratifying by Gender

GWAS analysis was also performed on UKBS imputed data stratified by gender using the '—filter-females' and '—filter-males' command in PLINK. Additive, linear association testing was then performed on female, and male data individually. 7,369,986 variants and 1358 individuals were included in male only analysis, of which 1333 had phenotypic data, and 7,373,492 variants and 1366 individuals were included in female only analysis, of which 1338 had phenotypic data.

4.3.4.1.2.2 Covariates

All GWAS analysis had a basic adjustment including four PCAs, panel, age, and gender as covariates (removing gender in gender-specific analysis). Full GWAS analysis also included absolute neutrophil and lymphocyte counts. Absolute monocyte and eosinophil counts data was not available for ALSPAC samples; therefore to keep analysis consistent between cohorts, MA and EA were not included as covariates for UKBS analysis.

4.3.4.2 Cohort Three - ALSPAC

4.3.4.2.1 Post-imputation

GWAS analysis was conducted on genotyping data which was imputed on the GRCh37-hg19 genome build to the HRC version 1.0 reference panel (McCarthy *et al.*, 2016). Imputed variants with a MAF <0.01 and imputation score <0.80 were removed, and then linear additive association tests were performed using logged, z-scored mtDNA CN as a continuous phenotype. 7,401,334 variants and 5461 mothers passed QC (Table 4.2). As a replication study, GWS threshold was set at *p* <1.00E-05.

4.3.4.2.1.1 Stratifying by DNA Source

Anna Guyatt at University of Bristol reported significant differences in mtDNA CN between samples which had been extracted from whole blood, or white blood cell pellets, therefore ALSPAC mothers were stratified by DNA source. 2056 ALSPAC mothers were included in whole blood analysis, and 3405 mothers were included in

white cell analysis. Basic and full adjustment GWAS were performed on all ALSPAC mothers, and then mothers stratified by DNA source.

4.3.4.2.1.2 Covariates

Basic adjustment included age, DNA concentration (ng/μL), and two PCAs as covariates. Source of DNA (white blood cell pellets, or whole blood cells) was also included in basic analysis when analysing all mothers (n=5461). Full adjustment analysis also included estimated neutrophil and lymphocyte cell counts as covariates.

4.3.5 Meta-analysis

PLINK linear association output files do not contain all the information needed to perform a meta-analysis, therefore UKBS GWAS summary data was manually manipulated in to the correct format before conducting meta-analysis, adding in imputation score and reference allele information (Figure 4.8). Variants in PLINK formatted, linear association output files were matched to their previously extracted imputation scores (details in section 4.3.3.1.3.2) using their chromosome and base position (Figure 4.8B). Columns were then reordered into the correct META format using ‘awk’ (Figure 4.8C). The final formatted files were then used for meta-analysis.

A

CHR	SNP	BP	A1	TEST	NMISS	BETA	SE	L95	U95	STAT	P
22	.	16050822	A	ADD	2671	0.03238	0.0353	-0.0368	0.1016	0.9175	0.359
22	rs715549	16249098	T	ADD	2671	-0.1129	0.06144	-0.2333	0.007553	-1.837	0.06632
22	rs76702808	16347540	A	ADD	2671	-0.1025	0.04282	-0.1864	-0.01857	-2.394	0.01675
22	rs2106719	16349432	A	ADD	2671	-0.08544	0.04254	-0.1688	-0.002059	-2.008	0.04471
22	rs2158040	16349564	A	ADD	2671	-0.0395	0.03521	-0.1085	0.02951	-1.122	0.2621
22	rs9617216	16364324	T	ADD	2671	-0.05809	0.03892	-0.1344	0.01819	-1.493	0.1356
22	rs5771604	16364923	A	ADD	2671	-0.02015	0.02866	-0.07632	0.03603	-0.7029	0.4822
22	rs9617220	16372719	T	ADD	2671	0.1093	0.07918	-0.04591	0.2645	1.38	0.1676
22	rs138132527	16393312	A	ADD	2671	-0.04544	0.03796	-0.1198	0.02896	-1.197	0.2314

B

CHR	SNP	BP	A1	TEST	NMISS	BETA	SE	L95	U95	STAT	P	REF	ALT	INFO
22	.	16050822	A	ADD	2671	0.03008	0.03524	-0.039	0.09915	0.8534	0.3935	G	A	0.804095
22	rs715549	16249098	T	ADD	2671	-0.1174	0.06131	-0.2376	0.002751	-1.915	0.05559	C	T	0.887087
22	rs76702808	16347540	A	ADD	2671	-0.1037	0.04275	-0.1875	-0.01995	-2.427	0.01531	T	A	0.857391
22	rs2106719	16349432	A	ADD	2671	-0.08649	0.04245	-0.1697	-0.00329	-2.037	0.0417	G	A	0.845634
22	rs2158040	16349564	A	ADD	2671	-0.03763	0.03514	-0.1065	0.03124	-1.071	0.2843	G	A	0.803467
22	rs9617216	16364324	T	ADD	2671	-0.05925	0.03887	-0.1354	0.01693	-1.524	0.1276	C	T	0.916923
22	rs5771604	16364923	A	ADD	2671	-0.0183	0.02863	-0.0744	0.03781	-0.6392	0.5228	C	A	0.852095
22	rs9617220	16372719	T	ADD	2671	0.1047	0.07912	-0.05035	0.2598	1.324	0.1858	G	T	0.832244
22	rs138132527	16393312	A	ADD	2671	-0.04691	0.0379	-0.1212	0.02738	-1.238	0.216	G	A	0.954365

C

chr	rsid	pos	allele_A	allele_B	test	nmiss	beta	se	L95	U95	stat	P_value	info
22	.	16050822	G	A	ADD	2671	0.03008	0.03524	-0.039	0.09915	0.8534	0.3935	0.804095
22	rs715549	16249098	C	T	ADD	2671	-0.1174	0.06131	-0.2376	0.00275	-1.915	0.05559	0.887087
22	rs76702808	16347540	T	A	ADD	2671	-0.1037	0.04275	-0.1875	-0.02	-2.427	0.01531	0.857391
22	rs2106719	16349432	G	A	ADD	2671	-0.0865	0.04245	-0.1697	-0.0033	-2.037	0.0417	0.845634
22	rs2158040	16349564	G	A	ADD	2671	-0.0376	0.03514	-0.1065	0.03124	-1.071	0.2843	0.803467
22	rs9617216	16364324	C	T	ADD	2671	-0.0593	0.03887	-0.1354	0.01693	-1.524	0.1276	0.916923
22	rs5771604	16364923	C	A	ADD	2671	-0.0183	0.02863	-0.0744	0.03781	-0.6392	0.5228	0.852095
22	rs9617220	16372719	G	T	ADD	2671	0.1047	0.07912	-0.05035	0.2598	1.324	0.1858	0.832244

Figure 4.8: Reformatting process of converting PLINK linear association output files [A], in to META software formatted files [C], by first matching variants with their imputation score using their chromosome and base position [B].

To make cohorts comparable respective to DNA extraction source, meta-analysis was conducted using GWAS summary data from all UKBS samples (n=2671), and ALSPAC samples extracted from white blood cells (n=3405). A further meta-analysis was conducted on only UKBS females (n=1338), with ALSPAC white blood cell extracted samples to make cohorts comparable on a DNA, and gender basis. Both types of analysis were performed using basic adjusted summary data, and then fully adjusted summary data (including cell counts). Random-effects meta-analysis was performed using META version 1.7 (Marchini *et al.*, 2007; Liu *et al.*, 2010). Variants where then removed if their heterogeneity score was -1, and if the variant was only present in one cohort. 7,183,687 variants were included in the final meta-analysis of all UKBS, and ALSPAC white cell mothers with basic adjustments. 7,068,168 variants were included in the final meta-analysis of all UKBS, and ALSPAC white cell mothers with full adjustments. 7,142,717 variants were included in the final meta-analysis of UKBS females only, and ALSPAC white cell mothers with basic

adjustments, and 7,034,240 variants were included in the final meta-analysis of UKBS females only, and ALSPAC white cell mothers with full adjustments.

4.3.5.1 Locus Zoom

Variants with *p* <=1.00E-05 were extracted from meta-analysis summaries, and regions with three or more variants associated to one loci were plotted using LocusZoom (Pruim *et al.*, 2010). Meta-analysis summaries were restricted to the chromosome where the loci of interest was located. Individual files were then uploaded to LocusZoom, and the variant with the smallest p value was used as the reference variant. SNPs with a flanking size ± 400 Kb were also plotted, using genome build hg19/1000 Genomes Nov 2014 EUR as the reference build.

4.3.6 Software

4.3.6.1 Cohort One - UKBS

GWAS analysis was conducted using PLINK v1.90b3p 64-bit (2 Aug 2014) (Purcell *et al.*, 2007). Manhattan and quantile-quantile (QQ) plots were generated using R Studio version 3.0.1 (2013-05-16). Manhattan plots were generated using an updated version of 'qqman' R package, revised by Professor Heather Cordell at Newcastle University. Ensembl (Aken *et al.*, 2016; Herrero *et al.*, 2016), and GeneCards (Stelzer *et al.*, 2002; Weizmann Institute of Science, 2015) were used to annotate variants to their respective genes. META version 1.7.0 was used to meta-analyse results (Marchini *et al.*, 2007; de Bakker *et al.*, 2008; Liu *et al.*, 2010).

4.3.6.2 Cohort Three - ALSPAC

GWAS were conducted using SNPTEST version 2.5.0 (Marchini *et al.*, 2007). Manhattan and QQ plots were created using the 'qqman' R package in R 3.0.1 [http://cran.r-project.org/web/packages/qqman/]. ANNOVAR was used to perform gene annotations [http://annovar.openbioinformatics.org/en/latest/] (Wang *et al.*, 2010b), and the 'clusterBed' function of 'bedtools' v2 [https://github.com/arq5x/bedtools2] (Quinlan and Hall, 2010) was used to group borderline significant loci into clusters, according to whether they were within 1 Mb of each other. META version 1.7.065 was used to meta-analyse results (Marchini *et al.*, 2007; de Bakker *et al.*, 2008; Liu *et al.*, 2010). Regional association plots were produced with LocusZoom v1.3 (Pruim *et al.*, 2010), using the 1000G March 2012 reference (the most recent hg19 source population available).

4.3.7 Statistical Power

Statistical power calculations were performed using a G×G or G×E interaction, Windows-based software program called QUANTO (Morrison and Gauderman, 2002), and were calculated on the assumption that mean mtDNA CN would be 200 copies/cell, with a standard deviation of 15% (equating to ±30 copies/cell) (further details in section 2.11).

4.3.7.1 Cohort One - UKBS

Given an estimated sample size of 3500 individuals, and a type one error rate of 5.00E-08 (expected GWAS significance level) (International HapMap, 2005; Pe'er *et al.*, 2008; Barsh *et al.*, 2012), power calculations indicate that we have >80% power to detect a significant association assuming a MAF >0.50, and an effect size (β_G) of >4.5 (Figure 4.9A). The power of the study will decrease as MAF or β_G decreases. If β_G is reduced to 4.0, then the study will only have 60% power for MAF >0.50. Similarly, if MAF >0.30, then the study will only have 65% power for a β_G of >4.5.

4.3.7.2 Cohort Three - ALSPAC

As a replicate cohort, we will attempt to replicate the top 1% variants (approximately 5000 variants) in the ALSPAC cohort, therefore power calculations were based on a type one error of 1.00E-05 instead. Power calculations were simulated on the same conditions as UKBS samples, but using 5461 ALSPAC samples available. There is 87.77% power to replicate nuclear variant with a MAF of 51%, but with a β_G of 3.2 (Figure 4.9B).

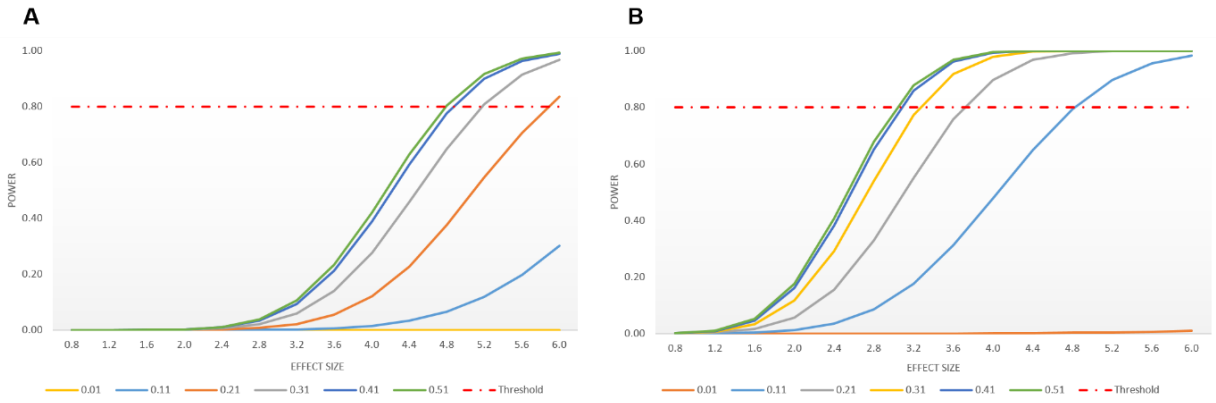


Figure 4.9: Simulated power calculations to determine a significant nuclear variant association with mitochondrial DNA copy number (μ=200 copies/cell, ±30 copies) in [A] UKBS samples (n=3091), where p=5.00E-08, and [B] replicate cohort ALSPAC samples (n=5461), where p=1.00E-05. Individual coloured lines represent variant MAFs – indicated under x-axis. Red dotted line indicates 80% power threshold.

4.3.7.3 Meta-analysis

Given a combined sample size of approximately 9000 individuals, and a type one error rate of 5.00E-08 (expected GWAS significance level) (International HapMap, 2005; Pe'er *et al.*, 2008; Barsh *et al.*, 2012), power calculations indicate that we have >80% power to detect a significant association assuming a MAF >0.10, and an β_G of >4.8. If MAF were to increase, we have >80% power to detect MAF >0.40, and an β_G of >3.0 (Figure 4.10).

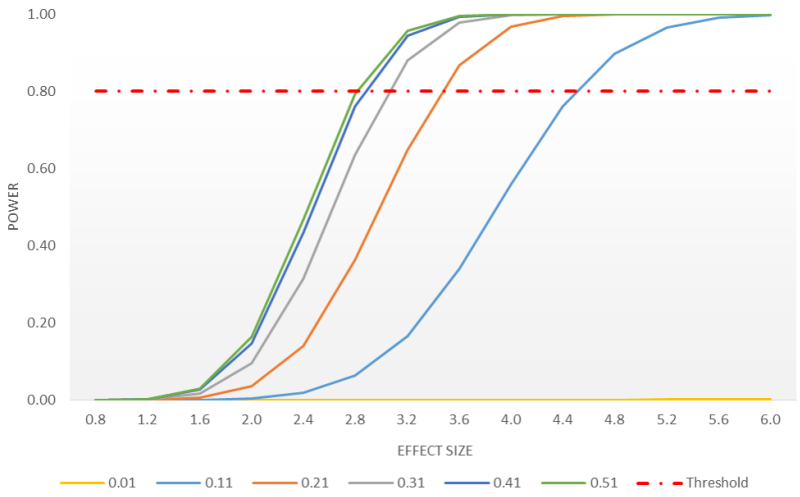


Figure 4.10: Simulated power calculation to determine a significant nuclear variant association with mitochondrial DNA copy number (μ=200 copies/cell ±30 copies) in a meta-analysis study of 9000 individuals, where p=1.00E-05. Individual lines represent variant MAFs. Red dotted line indicates 80% power threshold.

4.4 Results

4.4.1 Mitochondrial DNA Copy Number Distributions in Study Cohorts

To ensure mtDNA CN phenotypes were comparable between UKBS and ALSPAC cohorts, frequency distribution curves were generated for raw mtDNA CN and logged z-scored mtDNA CN in both cohorts (Figure 4.11). UKBS was also stratified by gender to visualise mtDNA CN distributions individually. As Figure 4.11 [A]-[D] show, the distributions of raw mtDNA CN are negatively skewed in both cohorts, however logging and z-scoring values normalised the distributions (Figure 4.11 [a]-[d]). Frequency distributions between UKBS gender cohorts were also comparable, therefore logged z-scored mtDNA CN values were used for all future analysis.

4.4.2 UKBS as a Discovery Study

4.4.2.1 GWAS using Non- Imputed Data

The UKBS cohort acted as a discovery cohort for this study. Additive linear QTL association tests were conducted using UKBS genotyping data (on the human NCBI36-hg18 build) which had been cleaned and restricted to autosomes. To identify covariates to include in association testing models, factors such as age, gender, and blood count information were adjusted for in individual analysis and compared. Association models selected were:

- Four PCAs and panel (UKBS samples were collected as two independent panels)
- Four PCAs, panel, and age
- Four PCAs, panel, and gender
- Four PCAs, panel, age, and gender
- Four PCAs, panel, age, gender, EA, MA, LA and NA
- Four PCAs, panel, age, gender, EA, MA, LA, NA, and PLT.

Four PCAs and panel were included in all analysis to adjust for population heterogeneity, and genotyping batch effects. GWAS results for each analysis are recorded in Figure 4.12, with respective QQ plots (Figure 4.13). Variants with a p value <=1.00E-06 were then identified and compared between analysis (Table 4.3).

As Figure 4.12 and Table 4.3 show, there is no difference in the top results when age, and gender are included in the GWAS model. This is also demonstrated by no

change in the lambda values in Figure 4.13a-d. Variant rs6421779 on chromosome 1 consistently has a genome-wide significant p value (p <5.00E-08) across all analysis. Variants rs1897023, rs2899527, rs12898299, and rs12901124 on chromosome 15 consistently have p values <=1.00E-06 across all analysis, except when blood count information is included as covariates. Variants rs10119358 on chromosome 9, and rs733631 on chromosome 16 are only associated (p <=1.00E-06) with mtDNA CN when blood count information is included in association models. rs6421779 on chromosome 1 is negatively associated with mtDNA CN, and all other variants are positively associated with mtDNA CN.

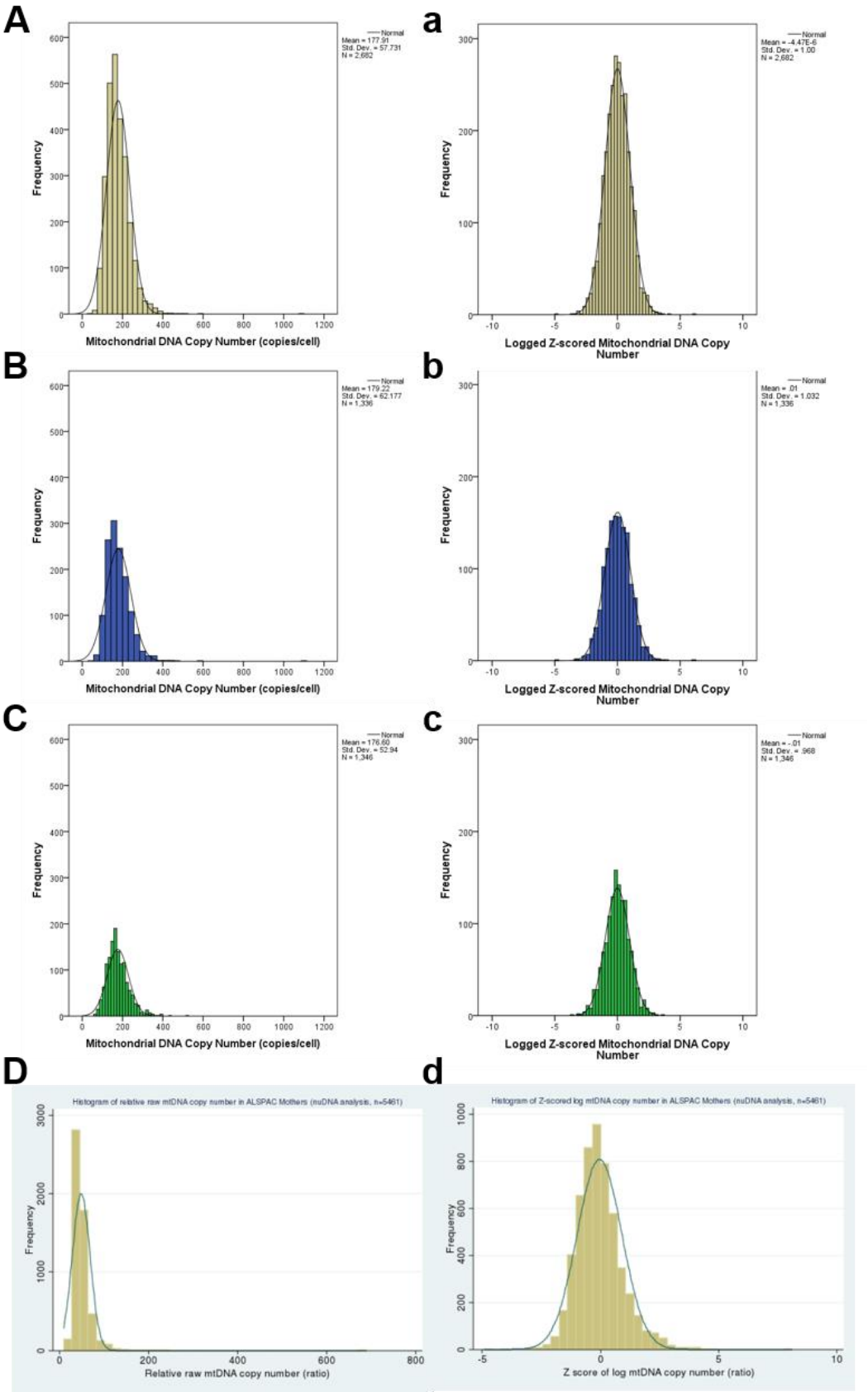


Figure 4.11: Frequency distribution plots of [A]-[D] mitochondrial DNA copy number (copies/cell) and [a]-[d] logged z-scored mitochondrial DNA copy number in UKBS and ALSPAC cohorts. [A, a] All UKBS samples (n=2682). [B, b] UKBS males only (n=1336). [C, c] UKBS females only (N=1346). [D, d] All ALSPAC samples (n=5461) – graphs provided by Anna Guyatt, Dr Santi Rodriguez's group, University of Bristol, UK.

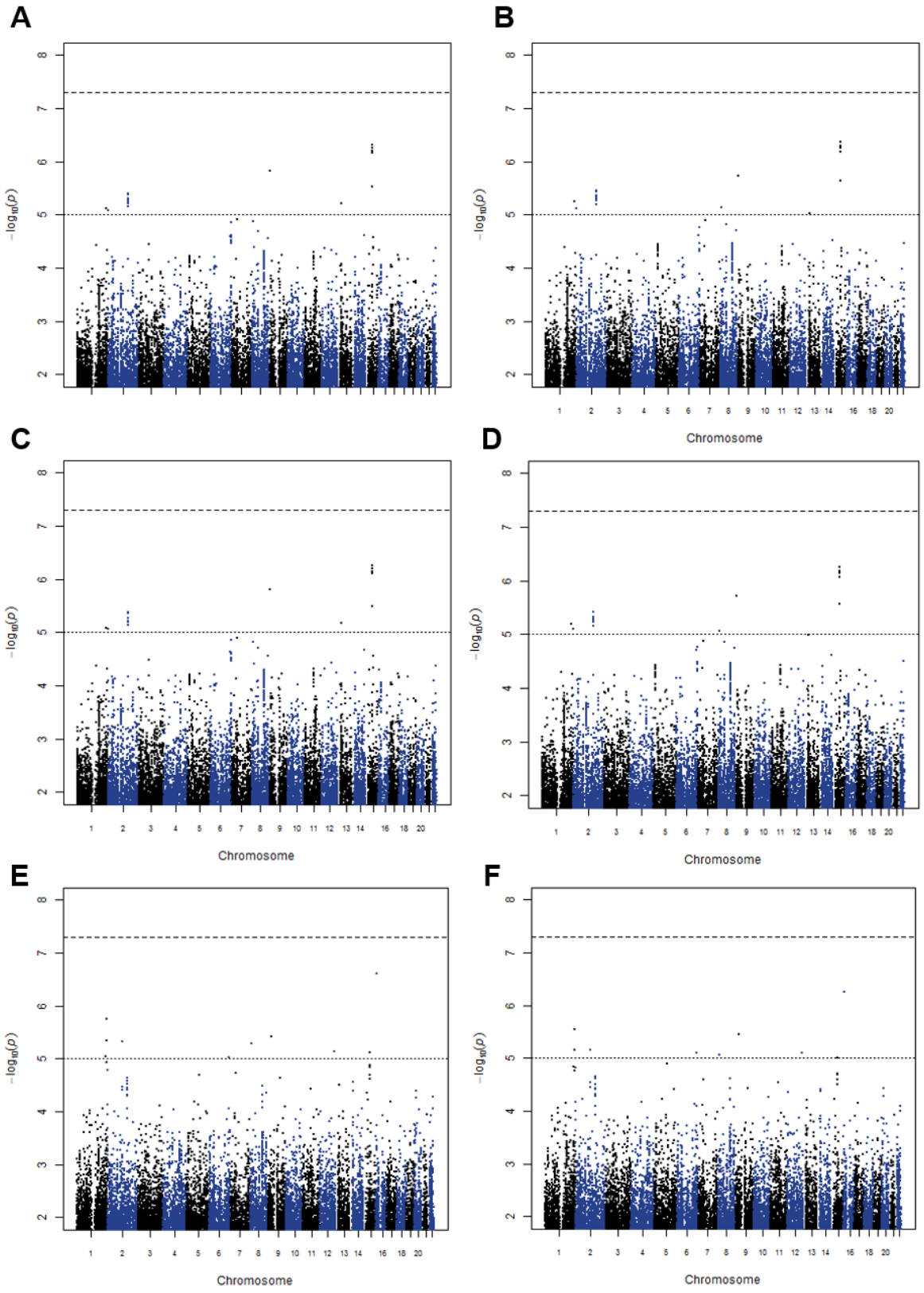


Figure 4.12: Manhattan plots of GWAS linear regression analysis of logged z-scored mitochondrial DNA copy number against 921,456 non-imputed genotyped nuclear SNPs in UKBS samples (n=2671). Tests adjusted for; [A] four PCAs and panel. [B] Four PCAs. panel, and age. [C] Four PCAs. panel, and gender. [D] Four PCAs. panel, age, and gender. [E] Four PCAs. panel, age, gender, and absolute eosinophil, lymphocyte, monocyte and neutrophil counts. [F] Four PCAs. panel, age, gender, absolute eosinophil, lymphocyte, monocyte and neutrophil counts, and platelet counts. Respective QQ plots are in Figure 4.13. Summary of top hits are in Table 4.3.

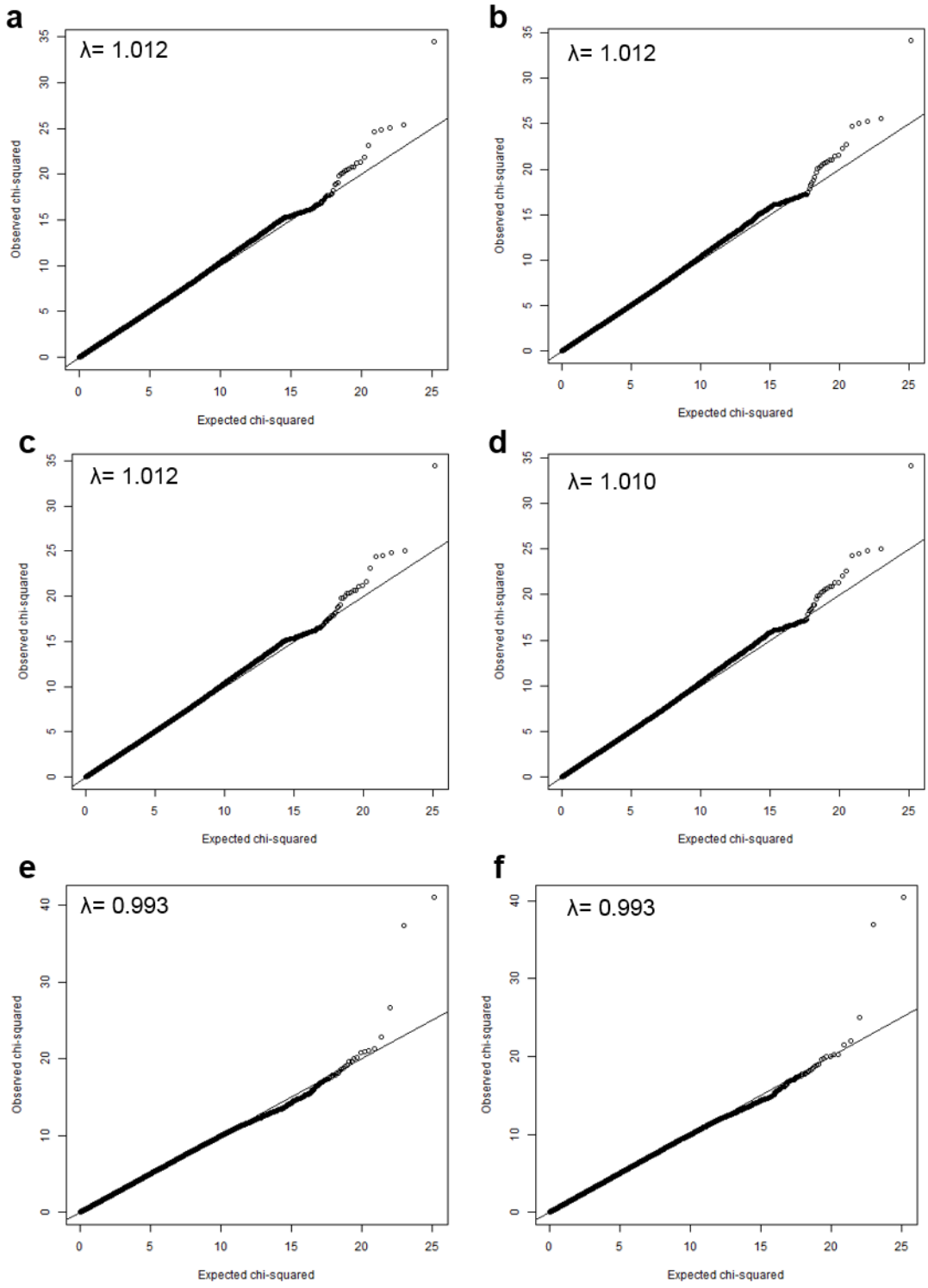


Figure 4.13: QQ plots of GWAS linear regression analysis of logged z-scored mitochondrial DNA copy number against 921,456 non-imputed genotyped nuclear SNPs in UKBS samples (n=2671). Tests adjusted for; [a] four PCAs and panel. [b] Four PCAs. panel, and age. [c] Four PCAs. panel, and gender. [d] Four PCAs. panel, age, and gender. [e] Four PCAs. panel, age, gender, and absolute eosinophil, lymphocyte, monocyte and neutrophil counts. [f] Four PCAs. panel, age, gender, absolute eosinophil, lymphocyte, monocyte and neutrophil counts, and platelet counts. X-axis: Expected chi-squared. Y-axis: Observed chi-squared. Respective Manhattan plots are in Figure 4.12.

						Four PCAs and Panel			Four PCAs, Panel and Age			Four PCAs, Panel and Gender			Four PCAs, Panel, Age and Gender			Four PCAs, Panel, Age, Gender, EA, LA, MA and NA			Four PCAs, Panel, Age, Gender, EA, LA, MA, NA and PLT			Associated Gene *(Neighbouring genes)*
CHR	*SNP ID*	*BP*	*A1*	*N*	*MAF*	*BETA*	*STAT*	*P*	*BETA*	*STAT*	*P*	*BETA*	*STAT*	*P*	*BETA*	*STAT*	*P*	*BETA*	*STAT*	*P*	*BETA*	*STAT*	*P*	
1	rs6421779	555887	G	2635	0.023	-0.5401	-5.888	4.42E-09	-0.5364	-5.860	5.21E-09	-0.5400	-5.886	4.45E-09	-0.5362	-5.858	5.28E-09	-0.6153	-6.435	1.51E-10	-0.6097	-6.385	2.08E-10	Non-coding (*MTCO3P12, MTND1P23, MTATP6P1, MTCO1P12, MTATP8P1, MTCO2P12, MTND2P28)*
15	rs1897023	52156136	G	2668	0.216	0.1674	5.046	4.82E-07	0.1678	5.069	4.28E-07	0.1666	5.015	5.64E-07	0.1663	5.017	5.60E-07							*UNC13c*
15	rs2899527	52176184	G	2667	0.206	0.1705	5.017	5.60E-07	0.1707	5.035	5.11E-07	0.1699	4.992	6.38E-07	0.1693	4.987	6.52E-07							*UNC13c*
15	rs12898299	52175765	A	2670	0.206	0.1694	4.996	6.23E-07	0.1697	5.017	5.60E-07	0.1686	4.967	7.22E-07	0.1681	4.966	7.28E-07							*UNC13c*
15	rs12901124	52173332	A	2669	0.215	0.1655	4.978	6.82E-07	0.1654	4.985	6.58E-07	0.1648	4.951	7.83E-07	0.1639	4.938	8.38E-07							*UNC13c*
9	rs10119358	493735	A	2213	0.053													0.4189	6.136	9.98E-10	0.4169	6.102	1.24E-09	*KANK1 (DMRT1, DOCK8)*
16	rs733631	4876092	A	2156	0.139													0.3928	5.176	2.48E-07	0.3817	5.018	5.66E-07	PPL (UBN1, SEC14L5)

Table 4.3: Details of top (p<=1.00E-06) non-imputed, genotyped variants which are linearly associated with logged, z-scored mitochondrial DNA copy number in UKBS samples (n=2671), when covariates such as PCAs, panel, age, gender and blood cell counts are adjusted for. PCA: Principal Components. EA: Absolute eosinophil counts. LA: Absolute lymphocyte counts. MA: Absolute monocyte counts. NA: Absolute neutrophil counts. PLT: Platelet counts. CHR: chromosome. BP: base position. A1: Minor allele. N: number of observations. MAF: Minor allele frequency. BETA: Regression coefficient. STAT: T-statistic. P: P-value.

4.4.2.2 GWAS using Imputed Data

UKBS samples were imputed using the Sanger Imputation Server (McCarthy *et al.*, 2016). Following imputation, linear GWAS tests were performed on all UKBS data (n=2671) using basic or full adjustments (Figure 4.14A and B respectively). Despite reporting no significant difference in mtDNA CN between genders (section 3.3.3), ALSPAC samples were all female, therefore UKBS samples were stratified by gender to maintain consistency between analysis, and identify if there were gender specific genetic associations with mtDNA CN differences. Basic and full adjustment linear GWAS analysis were performed on each UKBS group (Figure 4.14), and variants with a p value <=1.00E-06 were recorded (Table 4.4). Respective QQ plots are displayed in Figure 4.15.

As Figure 4.14 shows, GWAS results were different between genders. As Table 4.4 summarises, 24 variants were reported in all UKBS, basic adjusted analysis. 20 variants were located in gene *UNC13C* on chromosome 15, however only one variant was GWS (*p* <5.00E-08). This result was lost when cell count was adjusted for, and only three, single variants were reported in a long intergenic non-coding RNA (lincRNA), *KANK1*, and *PPL*. Variants rs75741190 in *KANK1*, and rs733631 in *PPL* were GWS. Three variants, which identified to one intergenic region, and *MYT1L* were top results for male only, basic adjusted analysis, however neither were GWS. Only rs13070764 in the intergenic region was replicated in full adjusted analysis, however one variant (rs77132217) in *SLC25A13* was GWS-associated with a positive change in mtDNA CN. In female only, basic analysis, only one top variant (rs733631) was reported to *PPL*. This was replicated when cell count was adjusted for, and was close to GWS (*p* =5.78E-08). Three variants in a lincRNA were also reported in this analysis, however they were not GWS.

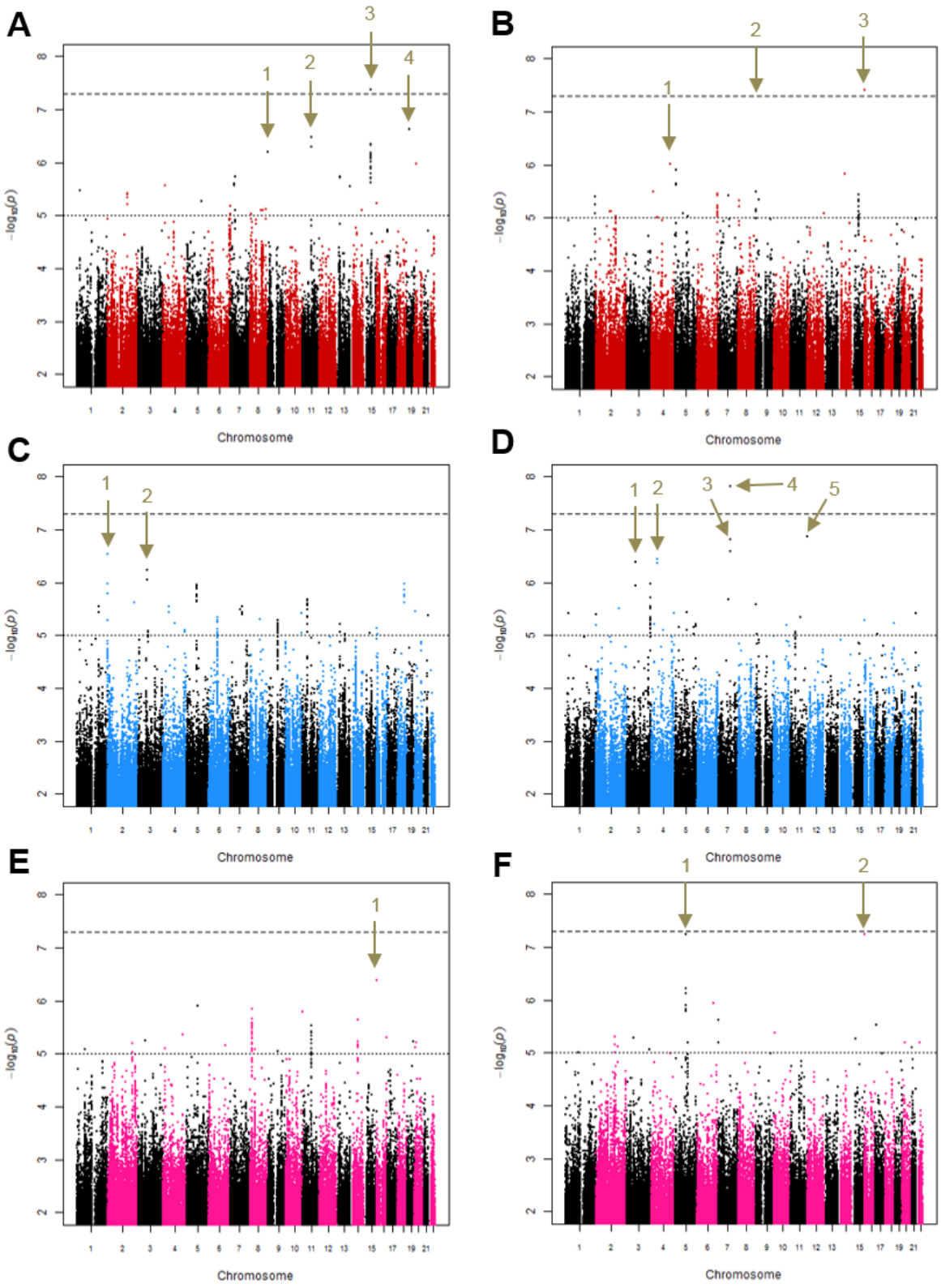


Figure 4.14: Manhattan plots of GWAS linear associations of imputed nuclear SNPs and logged z-scored mitochondrial DNA copy number in [A, B] all UKBS samples (n=2671), and UKBS samples stratified by [C, D] males (n=1333), and [E, F] females (n=1338). [A] Basic adjusted for four PCAs, panel, age and gender. [B] Fully adjusted for four PCAs, panel, age, gender, and absolute lymphocyte, and neutrophil counts. [C] and [E] Basic adjusted for four PCAs, panel, and age. [D] and [F] Fully adjusted for four PCAs, panel, age, and absolute lymphocyte, and neutrophil counts. Respective QQ plots are detailed in Figure 4.15. Arrows indicate variants where p<=1.00E-06 (summarised in Table 4.4).

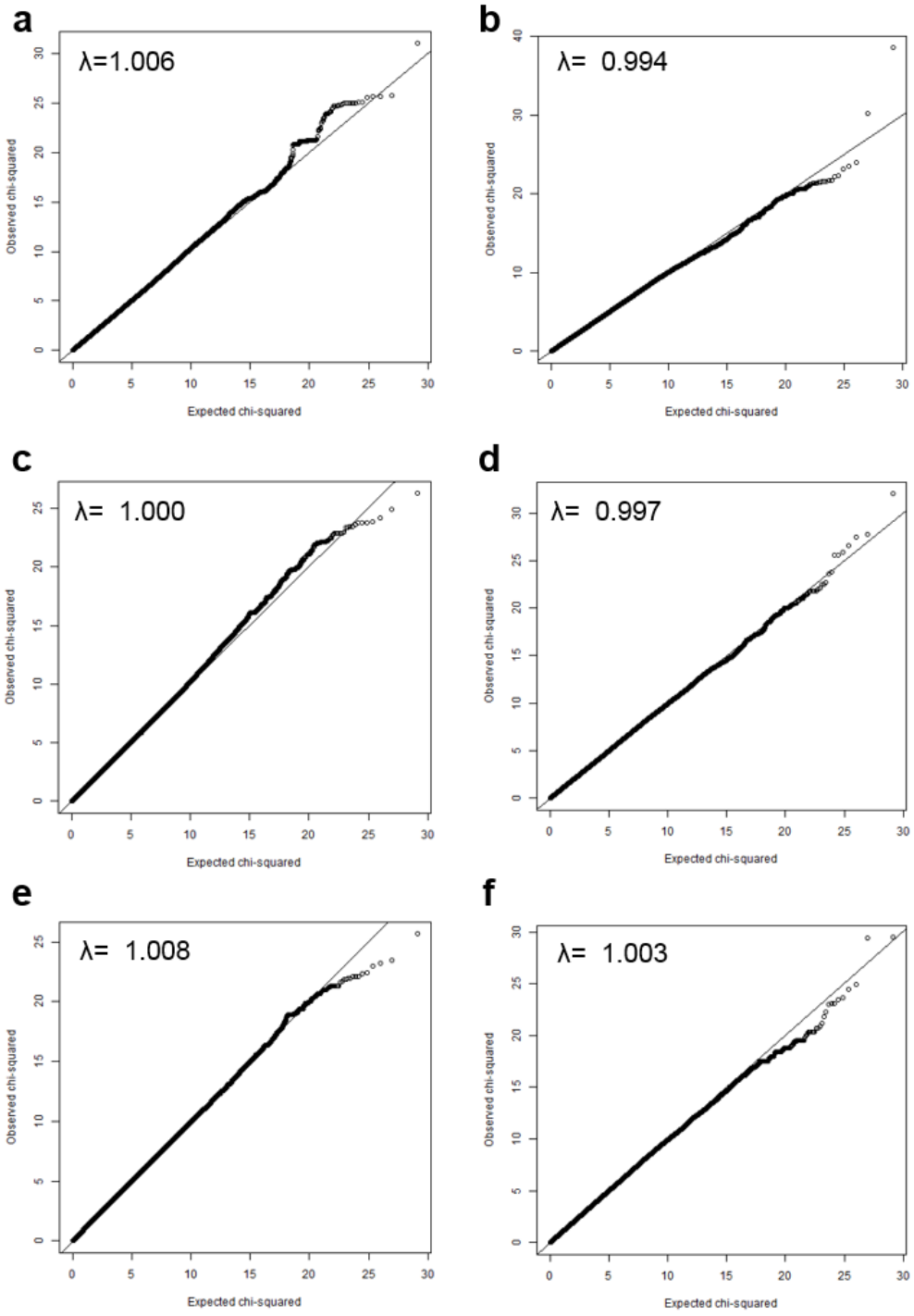


Figure 4.15: QQ plots of GWAS linear associations of imputed nuclear SNPs and logged z-scored mitochondrial DNA copy number in [a, b] all UKBS samples (n=2671), and UKBS samples stratified by [c, d] males (n=1333), and [e, f] females (n=1338). [a] Basic adjusted for four PCAs, panel, age and gender. [b] Fully adjusted for four PCAs, panel, age, gender, and absolute lymphocyte, and neutrophil counts. [c] and [e] Basic adjusted for four PCAs, panel, and age. [d] and [f] Fully adjusted for four PCAs, panel, age, and absolute lymphocyte, and neutrophil counts. X-axis: Expected chi-squared. Y-axis: Observed chi-squared. Respective Manhattan plots are in Figure 4.14.

Figure	Label	Chr	BP From	BP To	SNPs (N)	Average Beta	Average INFO	Max P	Gene
A	1	9	503184	503184	1	0.3227	0.9159	6.32E-07	*KANK1*
	2	11	68706854	68747529	2	0.4400	0.8820	3.31E-07	*IGHMBP2*
	3	15	54355939	54388892	20	0.1150	0.9960	4.20E-08	*UNC13C*
	4	19	13681163	13681163	1	0.5252	0.9017	2.34E-07	*CACNA1A*
B	1	4	150708460	150708460	1	-0.5820	0.8794	9.76E-07	RP11-526A4.1-001 LincRNA
	2	9	503184	503184	1	0.4523	0.9159	5.43E-10	*KANK1*
	3	16	4936091	4936091	1	-0.5820	0.8794	3.86E-08	*PPL*
C	1	2	1969704	1969704	1	0.4006	0.9845	2.94E-07	*MYT1L*
	2	3	73309903	73310337	2	-0.5100	0.8970	5.84E-07	Intergenic
D	1	3	73309903	73309903	1	-0.5085	0.8950	4.15E-07	Intergenic
	2	4	53605419	53627812	2	0.7160	0.9150	3.55E-07	Intergenic
	3	7	95652178	95660807	2	0.9890	0.9050	1.56E-07	*DYNC1/1*
	4	7	95807867	95807867	1	1.1070	0.9101	1.54E-08	*SLC25A13*
	5	11	131613807	131613807	1	-0.6123	0.8930	1.37E-07	*NTM*
E	1	16	4936091	4936091	1	0.5560	0.8433	4.06E-07	*PPL*
F	1	5	91979738	92000604	3	-0.4060	0.9760	5.70E-08	*RP11-133F8.2-003 LincRNA*
	2	16	4936091	4936091	1	0.6697	0.8433	5.78E-08	*PPL*

Table 4.4: Details of top variants (p <=1.00E-06) for linear association analysis of mitochondrial DNA copy number and imputed nuclear SNPs in UKBS samples, stratified by gender. Results reported summarise respective figure data in Figure 4.14.

4.4.3 ALSPAC as a Replication Study

4.4.3.1 Mitochondrial DNA Copy Number Assay Validation

Before performing genotype analysis between UKBS and ALSPAC cohorts, initial tests were performed to check that mtDNA CN values were not significantly different between UKBS and ALSPAC samples as a result of using different qPCR assays. Therefore, 185 randomly-selected, blinded UKBS samples were sent to Dr Rodriguez's group at University of Bristol, and 169 randomly-selected, blinded ALSPAC samples were sent to Newcastle University. PG, who originally measured mtDNA CN in all of the ALSPAC samples, no longer worked at University of Bristol therefore AG performed qPCR tests. Scatter plots were generated following qPCR analysis and compared (Figure 4.16).

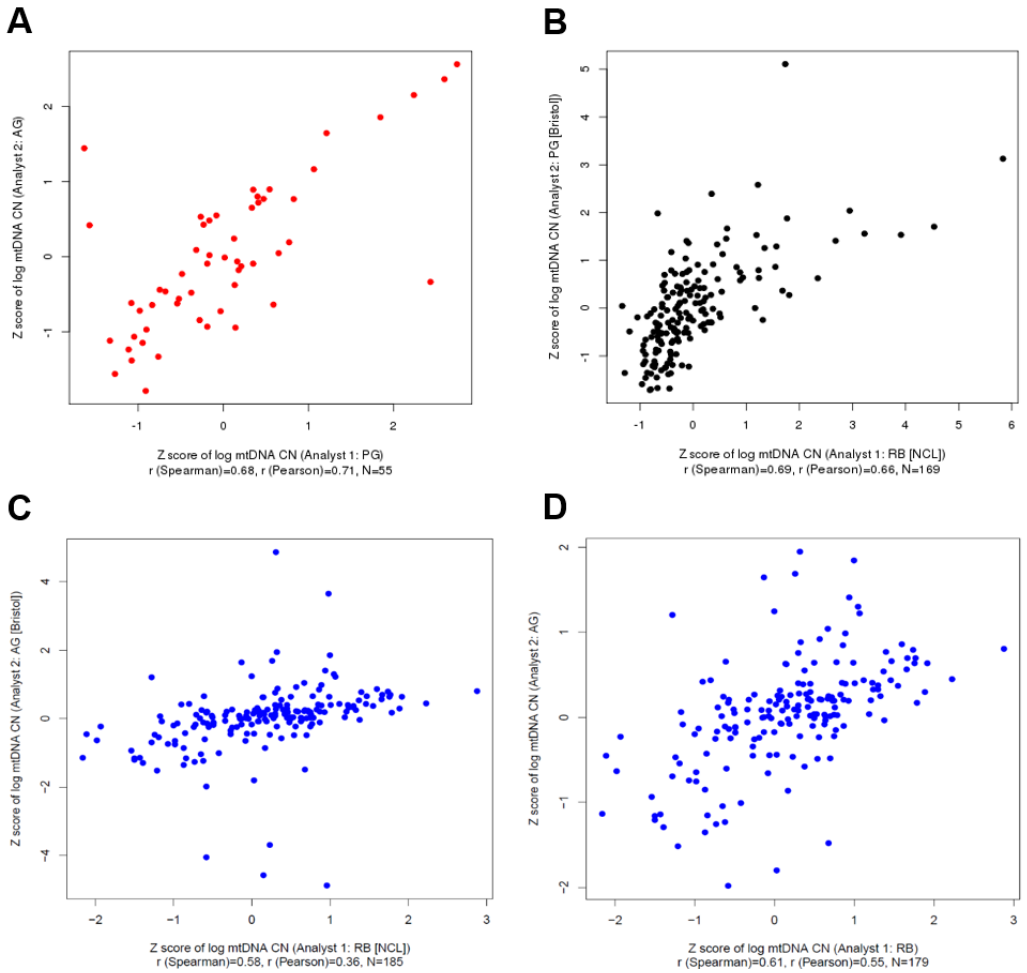


Figure 4.16: Scatter plots to identify the correlation of UKBS and ALSPAC Z-scored mitochondrial DNA copy number values when using different qPCR assays. [A] ALSPAC samples (n=55) measured by AG and PG at University of Bristol. [B] ALSPAC samples (n=169) measured by PG at University of Bristol, and RB at Newcastle University. [C] UKBS samples measured by RB at Newcastle University and AG at University of Bristol (n=185). Graphs generated by AG in Dr Santi Rodriguez's group at University of Bristol, UK.

As Figure 4.16A shows, qPCR measurements conducted on the same 55 ALSPAC samples by AG and PG generated mtDNA CN values which correlated with a Pearson's value of 0.71. This reassured that AG was able to produce comparable results to PG's work. ALSPAC samples, which were measured by RB at Newcastle University, and PG at University of Bristol, correlated with a Pearson's correlation of 0.66 (Figure 4.16B). UKBS samples measured by AG at University of Bristol correlated with measurements generated by RB at Newcastle University with a Pearson's value of 0.36 (Figure 4.16C), however this improved to 0.55 when six outlier samples were removed (±2 s.d) (Figure 4.16D). These results validated that mtDNA CN values between the cohorts were comparable, despite using different qPCR assays, however mtDNA CN values were logged and z-scored to normalise results to make them comparable between cohorts.

4.4.3.2 GWAS using Imputed Data

Linear association tests were performed using logged, z-scored mtDNA CN as a continuous trait. Analysis was performed on all ALSPAC mothers (n=5461), and then on mothers extracted from different blood sources – whole blood (n=2056), or white blood cell pellets (n=3405) (Figure 4.17). Samples were stratified because mtDNA CN differences were reported to be different when extracted from whole blood or white cell pellets (Hurtado-Roca *et al.*, 2016), and is dependent on blood cell composition (Pyle *et al.*, 2010), therefore we needed to adjust for this in our models. Basic adjustment and full adjustments (including cell counts) were performed on each group, and variants with a p value <=1.00E-06 were recorded (Table 4.5).

22 variants, with 19 variants identifying to one major loci spanning *PSMD3* on chromosome 17 in all ALSPAC, basic adjusted analysis. This signal was lost when cell counts were included in the analysis, however 63 variants had a p value <=1.00E-06, with 42 variants spanning 10 different loci (Table 4.5B). Analysis performed on ALSPAC mothers extracted from whole blood samples recorded five variants that were genome-significant, and three variants spanned the *TBC1D1* loci. These results were lost when cell counts were adjusted for in the model. No variants were genome-significant in whole blood, fully adjusted analysis (Table 4.5D). Analysis performed in ALSPAC mothers extracted from white blood cell pellets reported three variants - two variants were located in the *GPC6* loci on chromosome 19, and the other one in *BTBD2* on chromosome 13. These results were also not

replicated when cell counts were included in the association model. However, white blood cell, full adjusted analysis reported 113 variants with a p value <=1.00E-6, of which 80 variants associated to 18 loci (Table 4.5F). Between analysis models, only variants located in *ENPP2*, *EVL*, *PEX11G*, *RUNX1*, and downstream of *BACE2*, but upstream of *FAM3B* were replicated. These replications were only reported in fully adjusted analysis in all ALSPAC mothers, and white blood cell extracted mothers. All variants associated to loci regions were positively associated with mtDNA CN changes, except for associations in *MED24, EVL,* and *BTBD2* which were negatively associated.

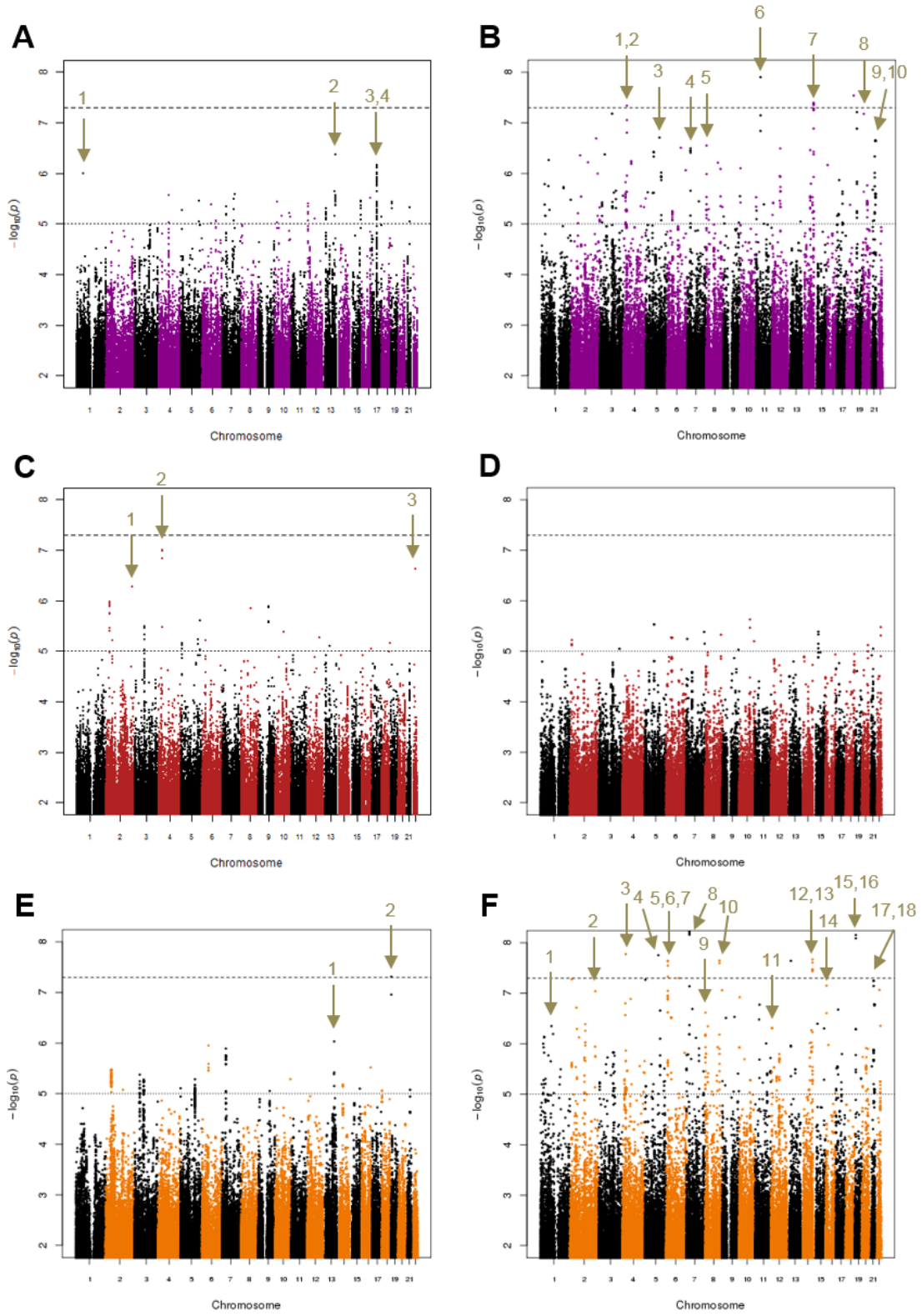


Figure 4.17: Manhattan plots of GWAS linear associations of imputed nuclear SNPs and logged z-scored mitochondrial DNA copy number in [A, B] all ALSPAC maternal samples (n=5461), and ALSPAC samples stratified by [C, D] samples extracted from whole blood (n= 2056), and [E, F] samples extracted from white blood cell pellets (n=3405). [A] Basic adjusted for two PCAs, DNA concentration, age, and DNA source. [B] Fully adjusted for two PCAs, DNA concentration, age, DNA source and estimated lymphocyte, and neutrophil counts. [C] and [E] Basic adjusted for two PCAs, DNA concentration, and age. [D] and [F] fully adjusted for two PCAs, DNA concentration, age, and estimated lymphocyte, and neutrophil counts. Respective QQ plots are detailed in Figure 4.18. Arrows indicate variants where p<=1.00E-06, and details are summarised in Table 4.5.

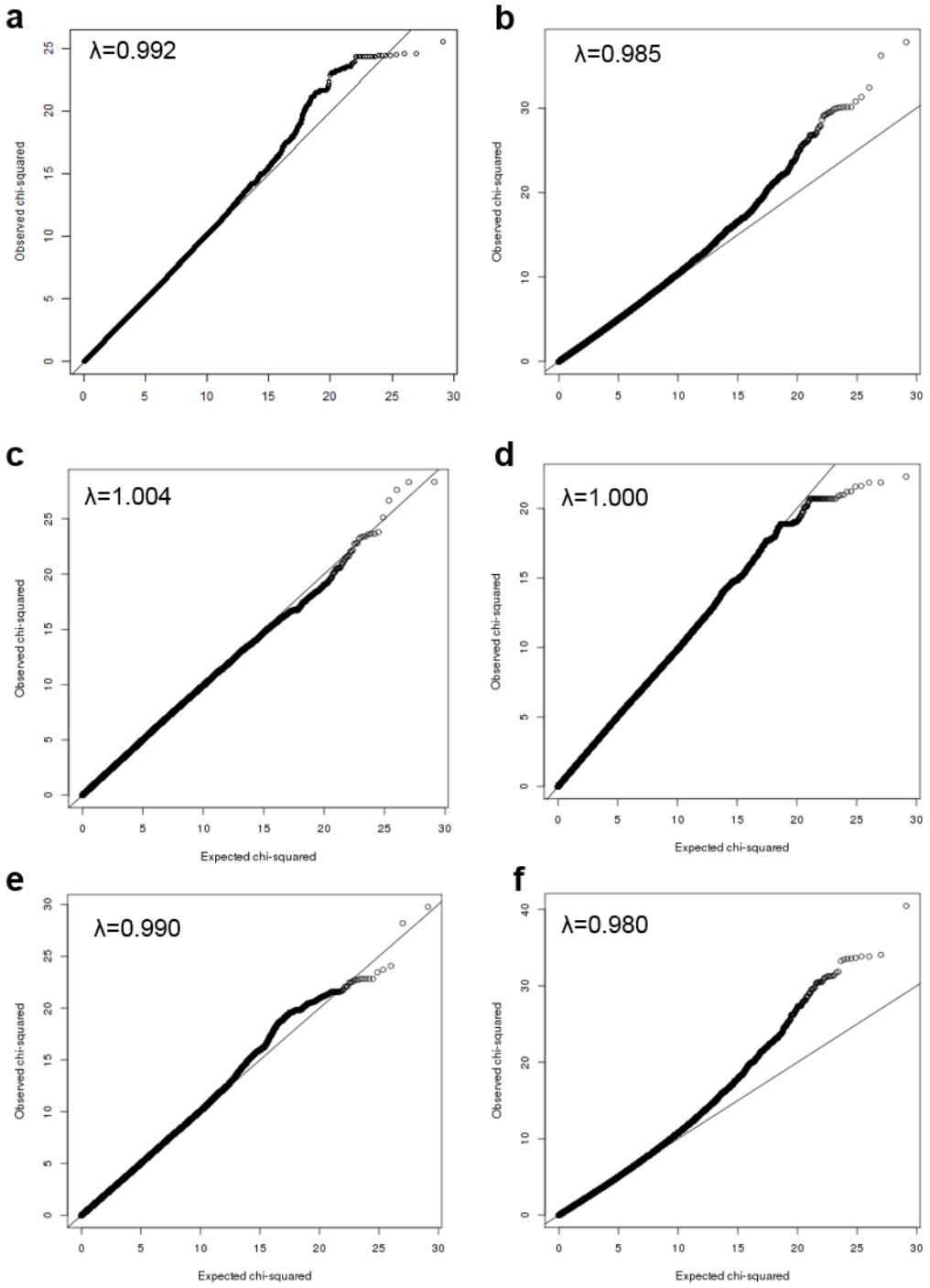


Figure 4.18: QQ plots of GWAS linear associations of imputed nuclear SNPs and logged z-scored mitochondrial DNA copy number in [A, B] all ALSPAC maternal samples (n=5461), and ALSPAC samples stratified by [C, D] samples extracted from whole blood (n= 2056), and [E, F] samples extracted from white blood cell pellets (n=3405). [a] Basic adjusted for two PCAs, DNA concentration, age, and DNA source. [b] Fully adjusted for two PCAs, DNA concentration, age, DNA source and estimated lymphocyte, and neutrophil counts. [c] and [e] Basic adjusted for two PCAs, DNA concentration, and age. [d] and [f] fully adjusted for two PCAs, DNA concentration, age, and estimated lymphocyte, and neutrophil counts. X-axis: Expected chi-squared. Y-axis: Observed chi-squared. Respective Manhattan plots are in Figure 4.17.

Figure	Label	Chr	BP From	BP To	SNPs (N)	Average Beta	Average INFO	Max P	Gene
A	1	1	55496648	55496648	1	0.1053	0.8968	9.95E-07	Intergenic
	2	13	104810437	104810437	1	0.1646	1.0000	4.34E-07	Intergenic
	3	17	38140927	38153554	19	0.0850	1.0000	6.94E-07	*PSMD3*
	4	17	38178627	38178627	1	-0.0863	0.9967	9.86E-07	*MED24*
B	1	4	30525905	30596920	3	1.5040	0.8920	4.57E-08	Intergenic
	2	4	68337581	68398591	3	1.2490	0.9510	5.87E-07	*CENPC*
	3	5	66961825	66979002	3	0.8550	0.9090	3.73E-07	RP11-434D9.1-002 LincRNA
	4	7	26557356	26562766	4	0.7060	0.9960	3.20E-07	Regulatory Region
	5	8	120600345	120626542	2	1.3850	0.9070	1.78E-09	*ENPP2*
	6	11	36627561	36810434	3	1.2980	0.8970	1.25E-08	*C11orf74*
	7	14	100577541	100635222	12	-0.3400	0.9970	4.06E-08	*EVL, DEGS2*
	8	19	7545964	7556451	2	0.5330	0.9020	6.14E-08	*PEX11G*
	9	21	36447535	36449047	2	1.0290	0.9080	4.61E-07	*RUNX1*
	10	21	42659146	42660973	8	0.8090	0.9980	2.24E-07	Downstream of *BACE2*, upstream of *FAM3B*
C	1	2	220890562	220890562	1	0.3651	0.8364	5.37E-07	Intergenic
	2	4	37902738	37904089	3	0.2200	0.9970	1.02E-07	*TBC1D1*
	3	22	34646160	34646160	1	0.5907	0.9536	2.41E-07	Intergenic
D									
E	1	13	93975643	93975643	1	-0.4035	0.8669	9.28E-07	*BTBD2*
	2	19	2011092	2011117	2	0.3470	0.8500	1.10E-07	*GPC6*
F	1	1	30087585	30094819	2	1.5360	0.9790	7.29E-07	Intergenic
	2	2	129328759	129384452	4	1.1700	0.9460	4.08E-07	Regulatory Region/ Intergenic
	3	4	26376117	26432449	2	2.0030	0.9950	1.58E-07	*RBPJ*
	4	5	129858500	130082868	2	1.7590	0.9770	2.20E-07	Intergenic
	5	6	14774456	14794512	13	1.4310	0.9790	2.27E-08	Intergenic
	6	6	43000197	43004551	3	2.1280	0.9940	3.02E-07	*MEA1*
	7	6	43023157	43038643	4	2.1260	0.9910	3.04E-07	*MRPL2*
	8	7	26557356	26607180	9	1.0480	0.9970	5.96E-09	Regulatory Region/ Intergenic
	9	8	2419872	2428673	6	0.8960	0.9900	2.41E-07	Intergenic
	10	8	120577056	120626542	2	2.3160	0.9310	2.01E-10	*ENPP2*
	11	12	8712321	8714192	3	0.7090	0.9650	4.89E-07	Intergenic
	12	14	91442756	91460300	2	1.8320	0.8940	5.40E-07	*RPS6KA5*
	13	14	100577541	100635222	10	-0.5260	0.9980	2.14E-08	*EVL*
	14	16	9456987	9465032	2	1.5940	0.9960	7.07E-08	RP11-243A14.1-001 LincRNA
	15	19	7545964	7556451	2	0.8430	0.8820	7.00E-09	*PEX11G*
	16	19	8032112	8049442	2	1.5450	0.9200	5.12E-07	*ELAVL1*
	17	21	36434894	36464229	4	1.5200	0.9460	5.38E-09	*RUNX1*
	18	21	42659146	42660973	8	1.3400	0.9960	1.66E-07	Downstream of *BACE2*, upstream of *FAM3B*

Table 4.5: Details of top loci with variants (p <=1.00E-06) for linear association analysis of mitochondrial DNA copy number and imputed nuclear SNPs in ALSPAC samples, stratified by DNA source. Results reported summarise respective figure data in Figure 4.17.

4.4.3.2.1 DNA Source between Cohorts

GWAS analysis conducted in whole blood extracted, and white blood cell extracted samples produced different results (Table 4.5). This suggested that DNA source affects the genetic associations of nuclear variants to mtDNA CN regulation. To keep analysis consistent between cohorts, and more genetically comparable, ALSPAC samples which were extracted from white blood cell pellets were used in meta-analysis against UKBS samples, which had also been extracted from white blood cell pellets.

4.4.3.2.2 Blood Count Data between Cohorts

mtDNA CN differs depending on blood cell composition (section 3.3.5), and GWAS analysis in ALSPAC samples adjusting for lymphocyte and neutrophil cell counts produced different results to analysis excluding cell count adjustments (Table 4.5). To keep analysis consistent between cohorts, and to adjust for any blood cell composition differences, meta-analysis will include cell count adjustments.

4.4.3.2.3 UKBS Replication

Variants which were GWS significant in UKBS analysis (p <5.00E-08) were identified in their respective white blood cell extracted ALSPAC analysis to investigate if they replicated as GWS (p <1.00E-05). The GWS hit in rs8039424 on chromosome 15, which identified as *UNC13C*, was not replicated in ALSPAC analysis (p= 4.58E-01, β= 0.020). The remaining 19 variants in *UNC13C* were also not GWS in ALSPAC analysis. The GWS variants rs75741190 and rs733631, on chromosomes 9 and 16 respectively, in fully adjusted all UKBS analysis were not present in ALSPAC data, and were therefore not replicated. In the UKBS male only, basic adjusted analysis, no variants were replicated in ALSPAC data. No variants were also replicated in the male only, fully adjusted analysis. The GWS rs77132217 variant (p= 1.54E-08), located on chromosome 7 to *SLC25A13,* p value was 9.89E-02 in ALSPAC analysis, and reported opposing beta values. The one variant reported in female only, basic analysis was not present in ALSPAC data, and was therefore not replicated. Two variants (rs7710144 and rs1154639), located on chromosome 5 to the lincRNA, were present in ALSPAC data, however they were not GWS, and reported opposing beta values.

Four GWS variants (p <=5.00E-08) were identified across the UKBS discovery analysis, with zero replication using the ALSPAC cohort. In addition, the ALSPAC

cohort identified over 200 GWS variants (*p* <=1.00E-05) across all analysis that were not seen in the UKBS cohort.

4.4.3.2.4 Nuclear Genotyping Data are Comparable at the SNP Level

To identify if the lack of replication in GWAS results was a result of differences in allele frequencies between the cohorts, MAFs were generated for both imputed and genotyped SNPs in both ALSPAC and UKBS cohorts. Variants were filtered to only include those that were present in both cohorts, leaving 7,156,393 SNPs which were then compared (Figure 4.19).

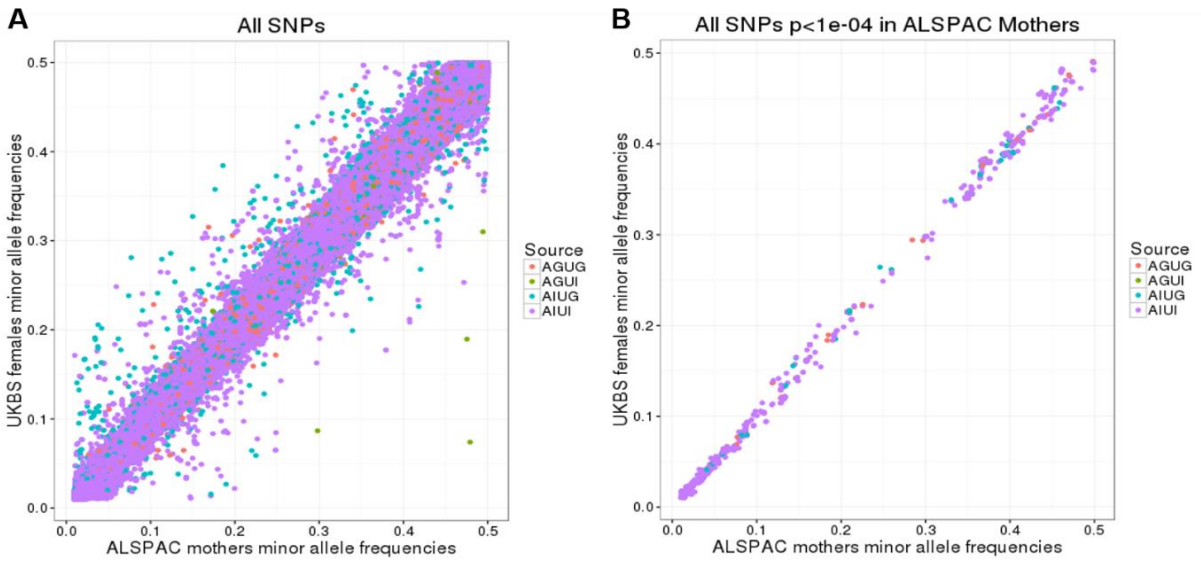


Figure 4.19: Scatter plots of MAFs of ALSPAC variants against UKBS variants. [A] All common variants between ALSPAC and UKBS (n=7,156,393). [B] Variants with a p value< 1.00E-04 in association tests. AGUG: ALSPAC genotyped, UKBS genotyped. AGUI: ALSPAC genotyped, UKBS imputed. AIUG: ALSPAC imputed, UKBS genotyped. AIUI: ALSPAC imputed, UKBS imputed. Graphs generated and provided by Anna Guyatt from Dr Santi Rodriguez's group at University of Bristol.

As Figure 4.19A shows, MAFs of variants in both cohorts were positively correlated, although there were several outliers. Analysis of variants with a *p* <1.00E-04 in GWAS analysis (Figure 4.19B) reported a stronger correlation with less outliers, suggesting MAFs were not affecting the top results reported between cohorts analysis.

4.4.4 Meta-analysis

Figure 4.19 demonstrates there was little difference in genotyping between ALSPAC and UKBS cohorts. However, Table 4.6 highlights other factors including mtDNA CN assay, mtDNA CN means, and blood count data contributing to heterogeneity between cohorts. Therefore, random effects meta-analysis was performed to adjust for population heterogeneity.

Bristol reported significant differences in mtDNA CN between samples extracted from whole blood cells, and samples extracted from white blood cell pellets (data not provided), and Figure 4.17 reports different genetic associations dependent on DNA source. Therefore to make meta-analysis as consistent, and comparable as possible, meta-analysis was performed only using ALSPAC samples extracted from white blood cells (n=3405).

As indicated by results in Figure 4.14 and Table 4.4, there may be gender-specific genetic effects contributing to mtDNA CN changes between males and females. As all ALSPAC samples were female, meta-analysis was performed using all UKBS samples (males and females) (n=2671), and then UKBS females only (n=1338).

As reported in section 4.4.3.2.2, blood cell composition also contributed to different GWAS associations to mtDNA CN changes. Therefore meta-analysis was conducted without cell count adjustments (basic adjustment), and with cell count adjustments (full adjustments).

Meta-analysis performed were:

- All UKBS and ALSPAC mothers extracted from white blood cells – basic adjustment
- All UKBS and ALSPAC mothers extracted from white blood cells – full adjustment
- UKBS females and ALSPAC mothers extracted from white blood cells – basic adjustment
- UKBS females and ALSPAC mothers extracted from white blood cells – full adjustment

Manhattan and QQ plots for all results are shown inFigure 4.20. Variants with a p value <=1.00E-06 were isolated from each set of results and are reported in Table 4.7.

As Table 4.7 shows, there were no GWS (*p* <5.00E-08) variants for any meta-analysis tests. Variant rs56325973, located in an intergenic region on chromosome 5, was strongly associated (*p*= 9.11E-07, β= 0.223) in all UKBS, fully adjusted meta-analysis with ALSPAC data, however it was not GWS, and was not replicated in other analysis. Variant rs116998574, located in *PLEKHB1* on chromosome 11, was also strongly associated (*p*= 8.83E-07, β= 0.444) in UKBS females, fully adjusted meta-analysis with ALSPAC data, however it was also not GWS or replicated across analysis.

	UKBS (Newcastle University)			ALSPAC (University of Bristol)		
	All	*Males*	*Females*	*All*	*Whole Blood*	*White Blood Cells*
Cohort Details	Unrelated individuals from across UK			Mothers only from former county of Avon, UK		
Original N	3091	1541	1550	5461	2056	3405
% Female		50			100	
Age Range (years)		17-69			16-43	
DNA Source		White Blood Cells		Mixed	Whole Blood	White Blood Cells
DNA Extraction Method		Phenol chloroform			Phenol chloroform	
qPCR Method		Multiplex *MT-ND1*, *MT-ND4*, and *B2M*			Singleplex mtDNA D-Loop (bases 317-381) and *B2M*	
qPCR Probes		TaqMan™			SYBR® Green	
Mean Copy Number * (+/- SD)	178.32 (56.46)	179.62 (59.43)	177.04 (53.33)	48.63 (19.78)		
Blood Count Data Source	Absolute proportions from Beckman- Coulter GenS			Estimated proportions from Illumina 450k array		
Blood Count	2617	1294	1323	546	245	301
Genotype + Phenotype Post-QC	2671	1333	1338	5461	2056	3405
Genotyping Chip		Illumina 1.2M			Illumina Human 660W HumanQuad	
Imputation Method		Sanger Imputation Server			Dr Gibran Hemani (University of Bristol), Dr George McMahon (University of Bristol), and Professor Jonathan Marchini (University of Oxford)	
Imputation Reference Panel		Haplotype Reference Consortium version 1.0				
Number of nDNA Variants Passed QC	7,441,490	7,369,986	7,373,492		7,401,334	

*Table 4.6: Demographic details of UKBS and ALSPAC cohorts. * indicates calculations based on individual s with genotyping and phenotype data (n=2671).*

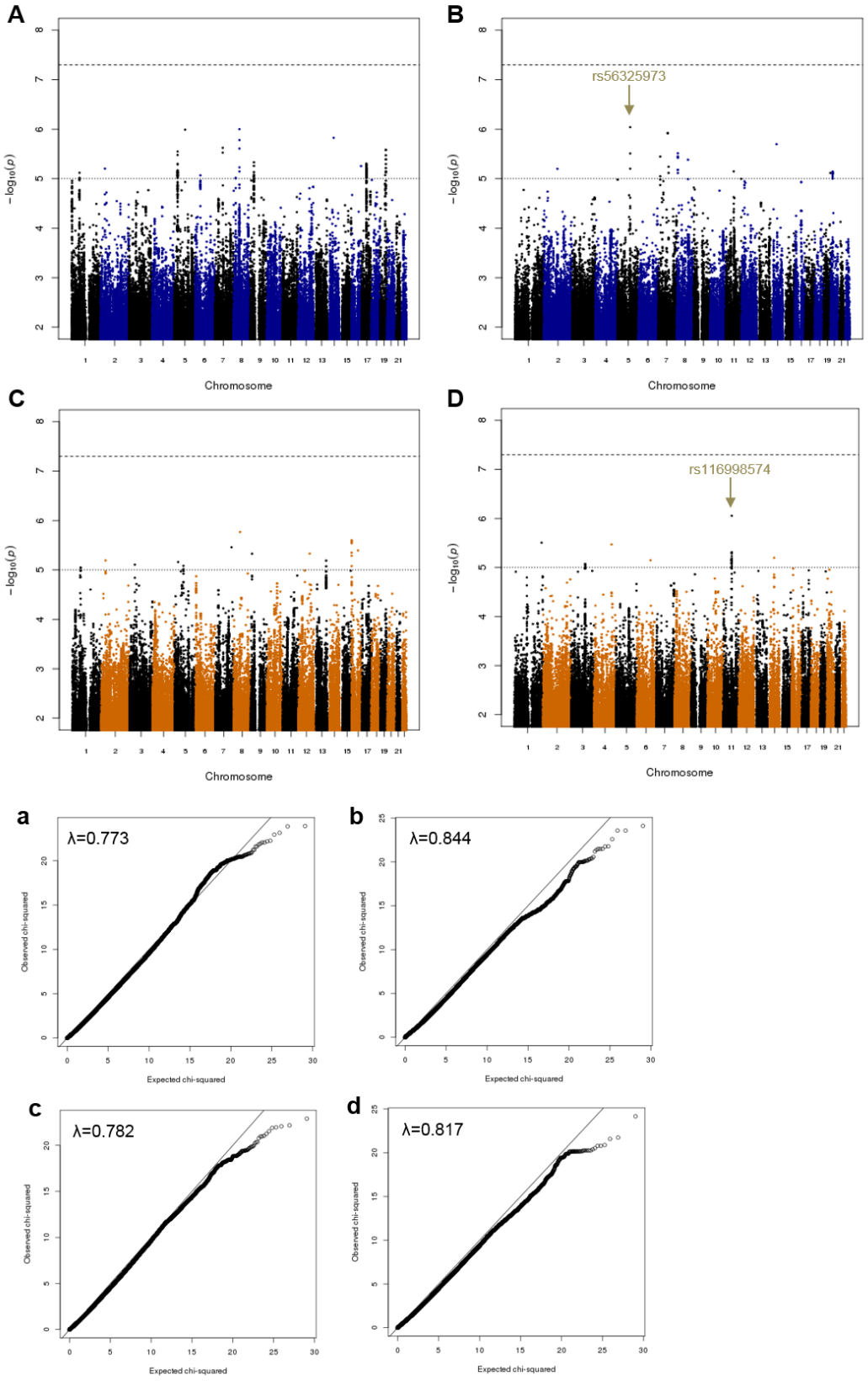


Figure 4.20: Manhattan [A]-[D] and QQ plots [a]-[d] showing the association of imputed nuclear variants with mitochondrial DNA copy number changes in UKBS and ALSPAC cohorts which had been meta-analysed using random effects. All UKBS (n=2671) and white blood cell extracted ALSPAC samples (n=3405) [A, a] not adjusted for cell counts, and [B, b] adjusted for cell counts. UKBS females (n=1338), and white blood cell extracted ALSPAC samples (n=3405) [C, c] not adjusted for cell counts, and [D, d] adjusted for cell counts. Arrows indicate variants with p<=1.00E-06 detailed in Table 4.7.

Figure	Chr	SNP ID	BP	Major Allele	Minor Allele	P value	Beta	SE	Q	Heterogeneity P value	I2	UKBS P value	ALSPAC P value	Gene
A														
Figure	Chr	SNP ID	BP	Major Allele	Minor Allele	P value	Beta	SE	Q	Heterogeneity P value	I2	UKBS P value	ALSPAC P value	Gene
B	5	rs56325973	109577493	T	C	9.11E-07	0.223	0.045	0.430	5.12E-01	0	9.55E-06	2.83E-02	Intergenic
Figure	Chr	SNP ID	BP	Major Allele	Minor Allele	P value	Beta	SE	Q	Heterogeneity P value	I2	UKBS P value	ALSPAC P value	Gene
C														
Figure	Chr	SNP ID	BP	Major Allele	Minor Allele	P value	Beta	SE	Q	Heterogeneity P value	I2	UKBS P value	ALSPAC P value	Gene
D	11	rs116998574	73359488	C	T	8.83E-07	0.444	0.090	0.266	6.06E-01	0	2.80E-05	9.99E-03	*PLEKHB1*

Table 4.7: Details of top variants (p <=1.00E-06) from meta-analysis performed between UKBS and ALSPAC datasets to identify nuclear variants associated with mitochondrial DNA copy number changes. Results correspond to their respective Manhattan plot in Figure 4.20.

4.4.4.1 Investigating Loci using Locus Zoom

Figure 4.20 showed there were some regions that may be associated with mtDNA CN, but are underpowered. Variants with *p* <1.00E-05 were extracted from each meta-analysis summary file, and regions with three or more variants in one loci were investigated, and LocusZoom plots were generated (Figure 4.21 and Figure 4.22).

In all UKBS, and ALSPAC basic adjusted analysis, 28 variants were identified in an intergenic region of chromosome 5 (maximum *p*= 2.82E-06, β=0.240) (Figure 4.21-1). Five variants were identified in *SNTG1* on chromosome 8 (maximum *p*= 1.00E-06, β= 0.091) (Figure 4.21-2), and 47 variants were identified in a loci spanning *CSF3, MED24, PSMD3,* and G*SDMA* on chromosome 17 (maximum *p*= 4.95E-06, β= -0.064) (Figure 4.21-3). These results were not replicated in full adjustment analysis, however three variants in another intergenic region on chromosome 5 were identified (maximum *p*= 9.11E-07, β= 0.214) (Figure 4.21-4). Seven variants spanning *CLDN23* on chromosome 8 (maximum *p*= 3.07E-06, β= -0.158) (Figure 4.21-5), and 13 variants were identified in *PCSK2* on chromosome 20 (maximum *p*= 7.22E-06, β= 0.169) (Figure 4.21-6).

In female only UKBS and ALSPAC analysis, three variants in loci *IQGAP2* on chromosome 5 were associated with mtDNA CN (maximum *p*= 8.26E-06, β= 0.037) (Figure 4.22-7), and seven variants in a domain spanning *NPRL3*, *POLR3K*, *SNRNP25*, *RHBDF1*, and *MPG* on chromosome 16 (maximum *p*= 2.50E-06, β= -0.140) (Figure 4.22-8). These results were not replicated in full adjustment analysis, however 41 variants in *PLEKHB1* on chromosome 11 were reported (maximum *p*= 8.83E-07, β= 0.360) (Figure 4.22-10), which was indicated as a top meta-analysis variant in Table 4.7. Four variants spanning a loci domain with *KPNA1, WDR5B, PARP9, DTX3L,* and *FAM162A* on chromosome 3 were also reported (maximum *p*= 8.50E-06, β= 0.159) (Figure 4.22-9).

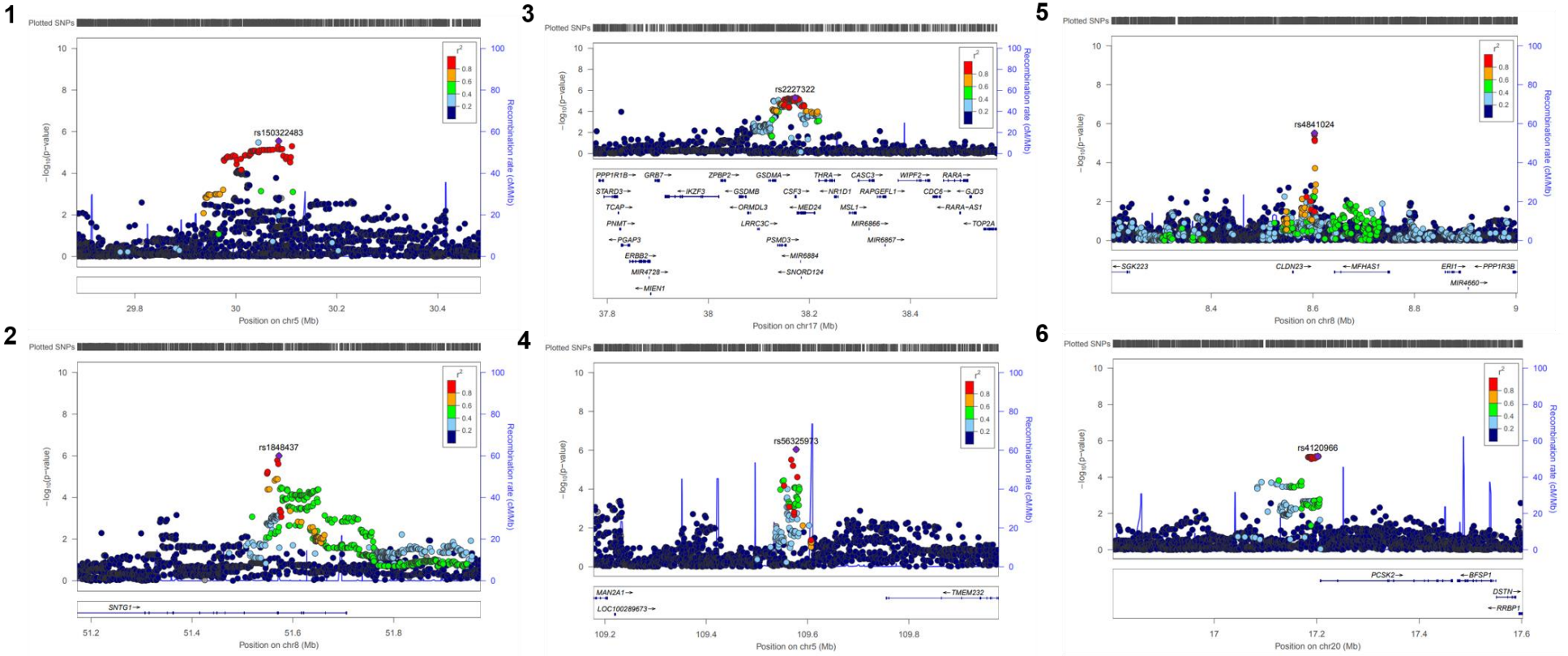


Figure 4.21: Locus Zoom plots for loci regions with three or more variants p<=1.00E-05 from meta-analysis conducted between all UKBS (n=2671) and ALSPAC females extracted from white blood cells (n=3405). Plots correspond to respective Manhattan plots in Figure 4.20 [A] and [B]. [1]-[3] Loci in basic adjusted meta-analysis in graph [A]. [4]-[6] Loci in fully adjusted meta-analysis in graph [B].

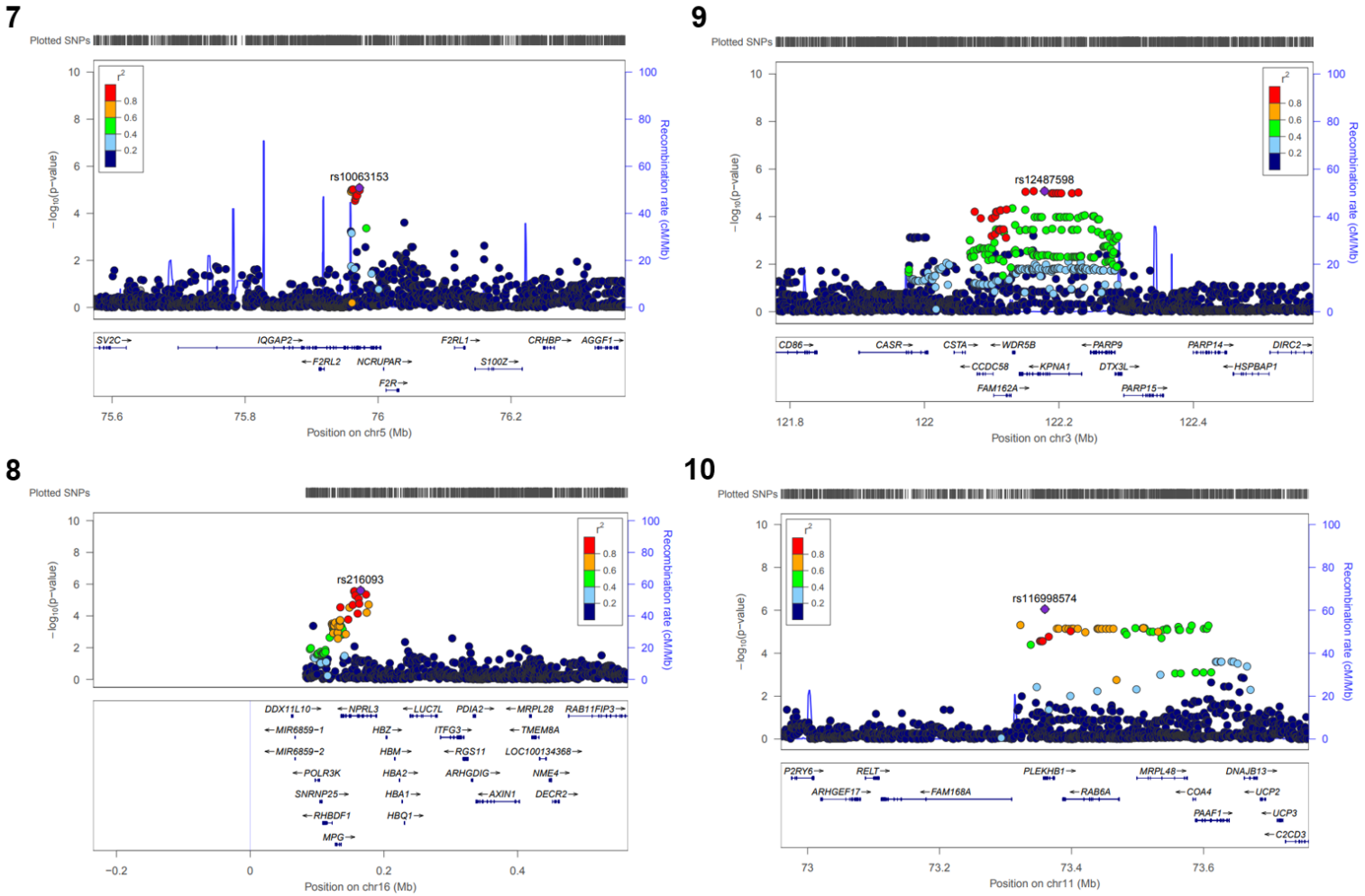


Figure 4.22: Locus Zoom plots for loci regions with three or more variants p<=1.00E-05 from meta-analysis conducted between UKBS females (n=1338) and ALSPAC females extracted from white blood cells (n=3405). Plots correspond to respective Manhattan plots in Figure 4.20 [C] and [D]. [7] and [8] loci in basic adjusted meta-analysis in graph [C]. [9] and [10] loci in fully adjusted meta-analysis in graph [D].

Figure	Locuz Zoom Image	Chr	Top SNP ID	Max P	SNPs in Loci (N)	Average Beta	SNPs Plotted	Genes
A	1	5	rs150322483	2.82E-06	28	0.240	3075	Intergenic
	2	8	rs1848437	1.00E-06	5	0.091	2315	*SNTG1*
	3	17	rs2227322	4.95E-06	47	-0.064	1512	*MED24, CSF3, GSDMA, PSMD3*
B	4	5	rs56325973	9.11E-07	3	0.214	1843	Intergenic
	5	8	rs4841024	3.07E-06	7	-0.158	3492	*CLDN23, MFHAS1*
	6	20	rs4120966	7.22E-06	13	0.169	2551	*PCSK2*
C	7	5	rs10063153	8.26E-06	3	0.037	2134	*IQGAP2*
	8	16	rs216093	2.50E-06	7	-0.140	1198	*NPRL3, POLR3K, SNRNP25, RHBDF1, MPG*
D	9	3	rs12487598	8.50E-06	4	0.159	2175	*KPNA1, WDR5B, PARP9, DTX3L, FAM162A*
	10	11	rs116998574	8.83E-07	41	0.360	1554	*PLEKHB1, RAB5A, MRPL48*

Table 4.8: Details of top SNPs from meta-analysis in Figure 4.20 which were used to generate Locus Zoom plots in Figure 4.21 and Figure 4.22, and their corresponding loci regions where p<=1.00E-05.

4.4.5 Mitochondrial Transcription Factor A (TFAM) Candidate Look Up

A candidate look up was conducted on the *TFAM* locus, which has been identified to be associated with mtDNA CN in one previous linkage study, and one GWAS which calculated a proxy mtDNA CN (Curran *et al.*, 2007; Cai *et al.*, 2015). Variants with MAF >0.01 and imputation quality >0.80 in and around *TFAM* (chr10: 60144782 - 60158981) were extracted, however no variants were significant (Figure 4.23).

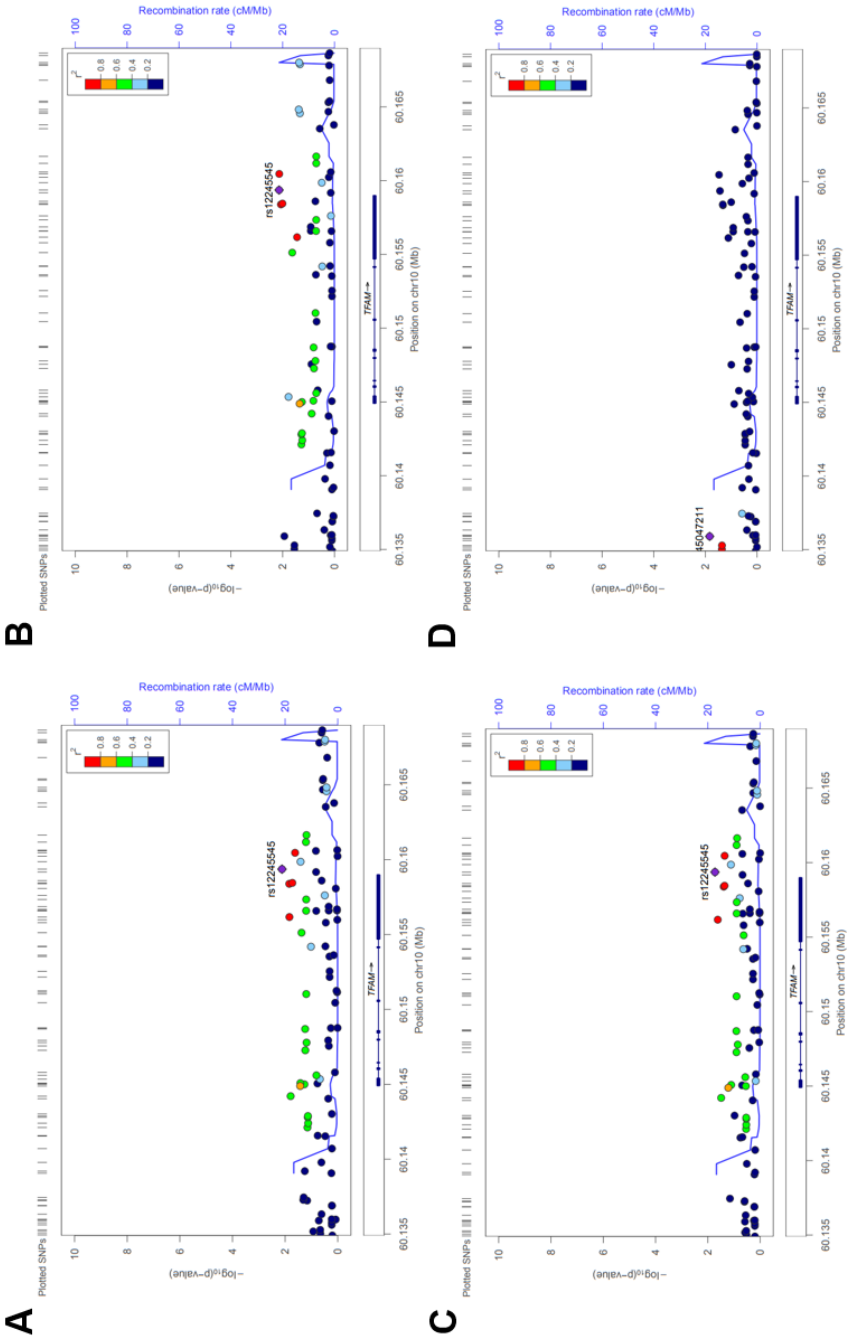


Figure	Analysis	Min P	Max P	SNPs Plotted
A	All UKBS, Female WC ALSPAC – Not Cell Adjusted	9.81E-01	7.59E-03	91
B	All UKBS, Female WC ALSPAC – Cell Adjusted	9.58E-01	7.28E-03	83
C	Female UKBS, Female WC ALSPAC – Not Cell Adjusted	9.97E-01	1.87E-02	91
D	Female UKBS, Female WC ALSPAC – Cell Adjusted	9.90E-01	1.46E-02	83

Figure 4.23: [A] – [D] Locus Zoom plots across Mitochondrial Transcription Factor A (TFAM) loci from meta-analysis conducted between UKBS and ALSPAC cohorts. [E] Table summarising p values across TFAM loci for each corresponding meta-analysis. WC: white cell count extracted samples.

4.5 Discussion

This is the largest, and most comprehensive study of nuclear genetic variants associated with directly-assayed mtDNA CN changes to date. Despite little genetic heterogeneity between UKBS and ALSPAC cohorts (Figure 4.19), our results report considerable differences in phenotypes and experimental measurements both between, and within both cohorts, and at the level of gender, DNA extraction source, and blood cell counts. The level of experimental heterogeneity reported may contribute to the lack of replication recorded between cohort analysis, resulting in no genetic loci being GWS (*p* <5.00E-08) after performing a meta-analysis. However, this study has identified critical variables that need to be considered, and adjusted for in future genetic studies.

4.5.1 Cohort One - UKBS Discovery

UKBS samples were a 50% mixture of male and females, and were all extracted from white blood cell pellets. Despite reporting no significant differences in mtDNA CN between genders, GWAS results suggest that there may be gender-specific genetic variables contributing to mtDNA CN regulation (Figure 4.14).

The discovery study conducted in the UKBS cohort reported four GWS nuclear variants associated to mtDNA CN changes across all analysis (Table 4.4 – section 4.4.2.2).

In the full study, unadjusted for cell count, 20 variants were located in *UNC13C* on chromosome 15, however only rs8039424 was GWS (*p*= 4.20E-08, β= 0.1150) (Table 4.4: Details of top variants (p <=1.00E-06) for linear association analysis of mitochondrial DNA copy number and imputed nuclear SNPs in UKBS samples, stratified by gender. Results reported summarise respective figure data in Figure 4.14.Table 4.4). *UNC13C* is a protein which is involved in vesicle maturation during exocytosis, and is a target of the diacylglycerol second messenger pathway. Although this protein has no direct links to mitochondrial biogenesis, the importance of efficient vesicle transportation has been identified to be important in mitochondrial dynamics, and regulating mitochondrial fusion and fission, as well as maintaining the architectural transmission within the mitochondrial inner membrane (Sugiura *et al.*, 2014). Emerging evidence is highlighting the importance of mitochondrial-derived vesicle (MDV) trafficking in immunological, and neurodegenerative diseases. For example, MDVs play a role in Parkinson's Disease development – reviewed by

(Roberts *et al.*, 2016), which has strong mtDNA CN associations (Pyle *et al.*, 2015a; Pyle *et al.*, 2015b).

In the full study, adjusted for cell count, analysis only one variant (rs75741190) in *KANK1* on chromosome 2 was GWS (*p*= 5.43E-10, β= 0.4523) (Table 4.4). This was also a top result for the full study, unadjusted for cell count analysis, however it was not GWS (*p*= 6.32E-07, β= 0.3227). *KANK1,* is a member of the Kank family proteins, which contain multiple ankyrin repeat domains*. KANK1* also has no direct associations with mitochondria, however knock out of *KANK1* leads to tumour development, and upregulation of *KANK1* results in increased apoptosis (Guo *et al.*, 2014). This suggests that *KANK1*'s role within regulating the cell cycle, may also be contributing to mitochondrial replication timings, as mtDNA CN has been reported to increase in cancer tumour development (Jiang *et al.*, 2005; Mambo *et al.*, 2005; Higuchi, 2007; Malik and Czajka, 2013). *KANK1* was replicated in cell count adjusted analysis, but only in the full UKBS cohort analysis, which suggested it may be a realistic candidate for mtDNA CN regulation.

In the full study, adjusted for cell count, analysis only one variant (rs733631) in *PPL* on chromosome 16 was GWS (*p*= 3.86E-08, β= -0.5820) (Table 4.4). This variant was replicated as a top result in both female-only analysis, but was not GWS*. PPL*, is a periplakin, which is a component of desmosomes. Desmosomes are intracellular junctions which provide adhesion between cells, but do not mediate intercellular or intracellular transport, and are known to interact with protein kinase-mediating signalling components, AKT1/PKB. Although no evidence suggests *PPL* has any direct effects on mitochondrial biogenesis, desmosomes and mitochondria are reported to form complexes under cellular stress, and can activate inflammatory responses (Rassat *et al.*, 1981). These events may be controlled by the AKT1/PKB pathway, which is reported to mediate signalling events which lead to inflammation or mitochondrial biogenesis, through activation of *PGC1-α,* and *NRF-1* and *-2* (Cherry and Piantadosi, 2015).

In male-only, cell count adjusted analysis, only one variant (rs77132217) in *SLC25A13* on chromosome 7 was GWS (*p*= 1.54E-08, β= 1.1070) (Table 4.4). *SLC25A13* on the other hand does have direct links to mitochondria, as it is a member of the mitochondrial carrier family. When stimulated by calcium, SLC25A13 catalyses the exchange of aspartate for glutamate and a proton across the inner

mitochondrial membrane (Thangaratnarajah *et al.*, 2014). Although no studies have investigated the direct effects of mitochondrial carrier proteins on mtDNA CN in humans, a study conducted in yeast identified a novel member of the mitochondrial carrier family located within a nDNA suppressor region which was vital for mtDNA metabolism and proper cell growth, and inactivation of this gene resulted in complete loss of mtDNA (Van Dyck *et al.*, 1995). This suggests that this gene may be contributing to mtDNA CN regulation, however the fact only one variant was identified to this region, and was not replicated in other analysis suggests this may be a false positive association.

No GWS results were reported in female-only analysis, however female-only, cell count adjusted analysis reported three variants in a loci of a lincRNA (*RP11-133F8.2*) on chromosome 5 which may be contributing to mtDNA CN regulation (maximum p= 5.70E-08, average β= -0.4060) (Table 4.4). This lincRNA does not have a known functional role. However, lincRNAs have been reported to be generated from mtDNA under regulation of nDNA, and they form intermolecular duplexes in cell- and tissue-specific quantities, which suggests they may have a functional role in the regulation of mitochondrial gene expression (Rackham *et al.*, 2011). These variants were not replicated in any other analysis, again suggesting they may be false positive results.

It is unclear why differences in results between males and females were reported in UKBS samples, because mtDNA CN was not significantly different between genders (p= 0.209 in section 3.3.3), and the MAFs of their variants were linearly correlated (Figure 4.24). Lopez and colleagues also reported differences in GWAS results by gender, despite no difference in mtDNA CN (López *et al.*, 2012). However, their study was performed in hundreds of individuals, which is underpowered for a GWAS. Future work would need to perform mtDNA CN studies in much larger cohorts, and stratify by gender to identify if genetic differences reported are a limitation of statistical power, or indeed are a result of different genetic associations to mtDNA CN regulation between genders.

Additionally, differences in mtDNA CN were reported between blood cell components, and blood cell composition was different between genders (section 3.3.5), which might suggest that blood composition is also contributing to the genetic differences reported between gender GWAS results.

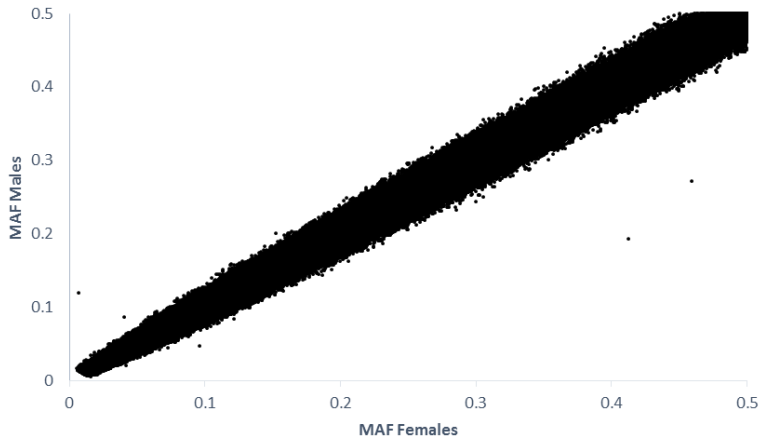


Figure 4.24: Minor allele frequencies of 921,456 genotyped variants in 1333 UKBS males and 1338 UKBS females.

4.5.2 Cohort Three - ALSPAC Replication

ALSPAC samples were only female, but were a combination of samples extracted from whole blood cells, and white blood cell pellets. This initially introduced a large amount of phenotypic variability to this study, because mtDNA CN measurements were significantly different when extracted from whole blood, or white blood cells (buffy coat) (Hurtado-Roca *et al.*, 2016). We adjusted for this as much as possible in this study, and limited our analysis to only include samples extracted from white blood cells, to make DNA source comparable to UKBS samples.

GWAS on all ALSPAC, and ALSPAC samples stratified by DNA source reported 37 different loci associated with mtDNA CN (Table 4.5 – section 4.4.3.2). Cell count adjusted analysis in all ALSPAC, and white blood cell extracted samples reported multiple variants that replicated in loci of *ENPP2*, *EVL*, *PEX11G*, and *RUNX1* (Table 4.5). However, no loci were replicated from UKBS analysis (Table 4.4 – section 4.4.2.2).

In the initial analysis, for all ALSPAC samples, with no cell count adjustment, 19 variants in the *PSMD3* loci on chromosome 17 were reported (p= 6.94E-07) (Table 4.5). *PSMD3* is a proteasome 26S subunit, non-ATPase 3 protein, and is associated with neutrophil count (Okada *et al.*, 2010). This association was removed when cell counts were introduced to the model (Figure 4.17B), which further suggests that mtDNA CN regulation is relative to blood cell composition, and may act as a proxy to cell count. This effect is also reported in females only, who have higher lymphocyte and neutrophil counts compared to males (section 3.3.5 - Table 3.3).

ENPP2 is a phosphodiesterase, and a phospholipase, which catalyses production of lysophosphatidic acid in extracellular fluids. Phosphodiesterases are reported to be important in mediating mitochondrial respiration, as they mediate cAMP signalling cascades in to the mitochondria, and promote mtDNA replication by stabilising TFAM (Acin-Perez *et al.*, 2011; Zhang *et al.*, 2015). Two variants were reported to be associated with *ENPP2*, with a maximum *p*= 2.01E-10, and β= 0.9310.

12 and 10 GWS variants respectively were associated with *EVL,* with a maximum *p*= 2.14E-08, and β= -0.5260 (Table 4.5). *EVL* is an actin-associated protein involved in a range of processes dependent on cytoskeleton remodelling and cell polarity. Mitochondria rely on the actin cytoskeleton for mobility, and to maintain their morphology through fusion and fission events (Boldogh and Pon, 2006). Fission and fusion events are vital for maintaining mitochondrial biogenesis, because they allow for mitochondrial replication machinery to be shared to maintain OXPHOS homeostasis and mitochondrial membrane potential.

Two variants were identified in *PEX11G,* with a maximum p value of 7.00E-09, β= 0.8820 (Table 4.5). *PEX11G* is a member of the PEX11 family which regulate the number and size of peroxisomes. As previously described (section 4.5.1), vesicle transport and peroxisome interactions with mitochondria are important because they rely on one another to maintain cellular metabolic homeostasis. Without efficient peroxisome-mitochondrial regulation, cells become stressed and can activate inflammasomes which can lead to a number of diseases (Demarquoy and Le Borgne, 2015).

Furthermore, four variants were reported to a loci of *RUNX1* (Table 4.5), a gene which is involved in the generation of hematopoietic stem cells. These variants had a stronger positive association with mtDNA CN than other variants (β= 1.5200, *p*= 5.38E-09), which suggests this loci may be important, especially as it was identified after cell count was adjusted for.

The associations in these genes have no direct effect on mtDNA CN, but similar to the associations reported in UKBS associated genes, there may be several genes involved in the regulation of mtDNA CN, and may all have differing levels dependent on the individual.

4.5.3 Meta-analysis

Meta-analysis between UKBS and ALSPAC cohorts reported two, independent variants in cell count adjusted analysis which identified an intergenic region on chromosome 5, and *PLEKHB1* on chromosome 11 respectively. However, both variants were not GWS (Table 4.7 – section 4.4.4).

PLEKHB1 is a pleckstrin homology domain containing B1 protein, and is associated with Wallerian Degeneration, and amyotrophic lateral sclerosis (ALS). Both Wallerian Degeneration and ALS are neurodegenerative diseases which develop as a result of immunological insults which are triggered from ROS accumulation (Dupuis *et al.*, 2004; Gaudet *et al.*, 2011). Mitochondrial biogenesis and mtDNA CN changes have been reported in several neurological-based diseases (Keeney and Bennett, 2010; Giordano *et al.*, 2014; Rice *et al.*, 2014; Pyle *et al.*, 2015a; Wei *et al.*, 2017), which may suggest that there are functional roles associated with these genes to overall mtDNA CN regulation.

As no genes were replicated in the meta-analysis, or across individual cohort GWAS, a list of possible candidate genes has been complied which were most significant, and consistent, between analyses' which would be beneficial to explore in future work (Table 4.9).

	Chr	Top SNP ID	Max P	Average Beta	Gene	Gene Name	Function	Associated Phenotypes	MitoCarta Score
UKBS only	5	rs7710144	5.70E-08	-0.4060	RP11-133F8.2-003 LincRNA	Uncharacterized LOC105379082	Unknown	None Reported	NA
	9	rs75741190	5.431E-10	0.4523	*KANK1*	KN Motif And Ankyrin Repeat Domains 1	Cytoskeleton formation by regulating actin polymerization	Cerebral Palsy, Spastic Quadriplegia 2, and Inherited Congenital Spastic Tetraplegia	NA
	15	rs8039424	4.20E-08	0.1150	*UNC13c*	Unc-13 Homolog C	Vesicle maturation during exocytosis as a target of the diacylglycerol second messenger pathway	None Reported	NA
	16	rs733631	3.86E-08	-0.5820	*PPL*	Periplakin	A component of desmosomes and the epidermal cornified envelope in keratinocytes. Is mediated by AKT1/PKB signalling, and is involved in the innate immune system and Butyrophilin (BTN) family interactions.	Paraneoplastic Pemphigus and Pemphigus	NA
ALSPAC only	7	rs7807582	5.96E-09	1.0480	Regulatory Region/ Intergenic	None	Unknown	None Reported	NA
	8	rs140416442	2.01E-10	2.3160	*ENPP2*	Ectonucleotide Pyrophosphatase/ Phosphodiesterase 2	Functions as both a phosphodiesterase, and a phospholipase, which catalyses production of lysophosphatidic acid (LPA) in extracellular fluids. Is also involved in the metabolism of water-soluble vitamins and cofactors and phospholipases	Teratocarcinoma and Intrahepatic Cholestasis	NA
	14	rs941898	2.14E-08	-0.5260	*EVL*	Enah/ Vasp-Like	Actin-associated proteins involved in a range of processes dependent on cytoskeleton remodeling and cell polarity. EVL enhances actin nucleation and polymerization, and related pathways are TCR signalling and Innate Immune System	Vein Disease	NA
	17	rs8066582	6.94E-07	0.0850	*PSMD3*	Proteasome 26S Subunit, Non-ATPase 3	Component of the complex, multiprotein 26S proteasome involved in the ATP-dependent degradation of ubiquitinated proteins. Maintains protein homeostasis by removing misfolded or damaged proteins. The proteasome participates in numerous cellular processes, including cell cycle progression, apoptosis, or DNA damage repair.	None Reported	NA
	17	rs3213762	9.86E-07	-0.0863	*MED24*	Mediator Complex Subunit 24	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Mediator serves as a scaffold for the assembly of RNA polymerase II and general transcription factors. Related pathways are PEDF Induced Signaling and Regulation of lipid metabolism by Peroxisome proliferator-activated receptor alpha (PPARalpha).	Transposition Of The Great Arteries	NA
	19	rs2335116	7.00E-09	0.8430	*PEX11G*	Peroxisomal Biogenesis Factor 11 Gamma	The protein encoded by this gene is a member of the PEX11 family which regulate the number and size of peroxisomes, and may induce clustering of peroxisomes.	None Reported	NA
	21	21:36449047_C_A	5.38E-09	1.5200	*RUNX1*	Runt Related Transcription Factor 1	RUNX1 is a transcription factor binds with cofactor CBFB to stabilise the RUNX1 protein. RUNX1 is involved in the generation of hematopoietic stem cells, and their differentiation into myeloid and lymphoid lines.	Platelet Disorder, Familial, With Associated Myeloid Malignancy and Acute Myeloid Leukemia With Translocation	NA
Meta-analysis	5	rs56325973	9.11E-07	0.2227	Intergenic	None	Unknown	None Reported	NA
	11	rs116998574	8.83E-07	0.4445	*PLEKHB1*	Pleckstrin Homology Domain Containing B1	Required for proper localization of retinogeniculate projections but not for eye-specific segregation	Wallerian Degeneration and Amyotrophic Lateral Sclerosis 1	NA

Table 4.9: Details of gene regions with most consistent, and significant variants associated with mitochondrial DNA copy number across all GWAS analysis. Gene details provided by GeneCards (http://www.genecards.org/) (Stelzer et al., 2002; Weizmann Institute of Science, 2015), and MitoCarta scores were generated using MitoCarta2 (Calvo et al., 2016).

4.5.4 Other Plausible Genetic Candidates

Somatic cells contain the same nDNA code, but different mtDNA CN. Several intergenic loci were reported as top variants in analysis in this study, which suggests there may be other epigenetic or biochemical factors contributing to differences in mtDNA CN.

nDNA is highly regulated by epigenetic modifications, and although mtDNA is not commonly known to be epigenetically modified, original studies of mtDNA do report mtDNA to contain between 12 to 36 methylcytosine residues per mtDNA molecule in certain animal cell lines, averaging at one 5'-mehtylcytosine per 500 cytosine residues in mtDNA compared to one 5'-mehtylcytosine per 36 cytosine residues in nDNA (Nass, 1973).

mtDNA CN levels are reported to be different between cancers (Malik and Czajka, 2013), and this may be a result of cells being in an undifferentiated or pseudo-differentiated state (Dickinson *et al.*, 2013), so when cancer cells try to undergo differentiation, they are unable to replicate their mtDNA in a regulated manner, which leads to inconsistent levels reported depending on the cancer type and stage (St. John, 2016). Treating cancer cells with DNA demethylation reagents, such as 5'-azacytidine or Vitamin C, results in a synchronous increase in mtDNA CN levels, and when DNA demethylation reagents are removed, cancer cells return to their 'pseudo-differentiated' state (Ebos *et al.*, 2009; Zuniga *et al.*, 2010), and resume their previous levels of mtDNA CN (Lee *et al.*, 2015; St. John, 2016). This suggests epigenetic regulation is overseeing mtDNA CN maintenance (Figure 4.25).

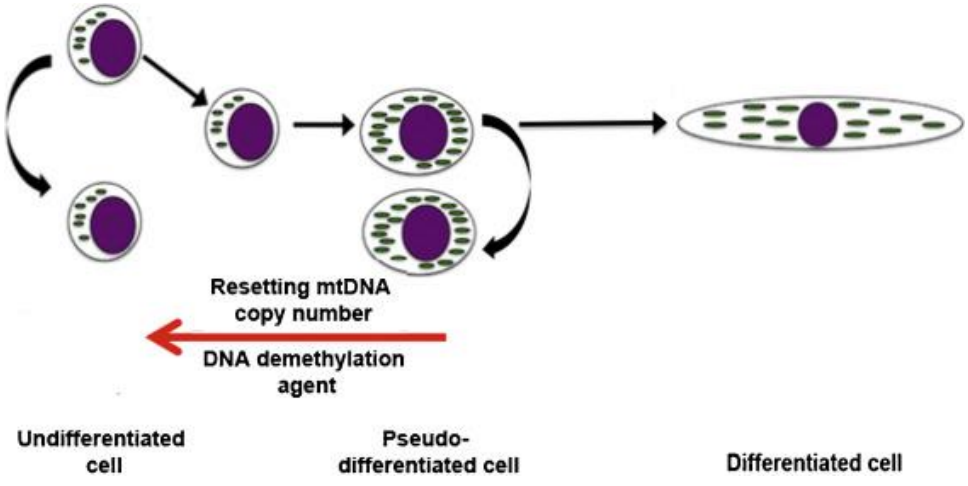


Figure 4.25: Mitochondrial DNA copy number changes when reprogramming cancer cells. Image from (St. John, 2016).

Kelly and colleagues also highlight the importance of nuclear-mitochondrial interactions in mitochondrial biogenesis regulation, because they report that reprogrammed pluripotent somatic cells are unable to control mtDNA CN levels or regulate ATP content, and this may be a result of *POLGA* and *PEO1* which were expressed at different levels (Kelly *et al.*, 2013b). This suggests mtDNA CN is not completely regulated by genetic code, but also epigenetic machinery and cell signalling.

Cell cycle timing is also critical for mitochondrial regulation (Bogenhagen and Clayton, 1977), and 80% of mtDNA is associated at polar ends of mitochondria during cell cycle (Nass, 1969). For successful pro-nuclear transfer, nuclei need to be transferred in early stages of the cell cycle to increase success rates. Hyslop and colleagues also showed that as a result of pro-nuclear transfer mitochondrial gene expression did not change (Hyslop *et al.*, 2016). Additionally, studies in mice have shown that mixing mtDNA and nDNA interactions across generations has no negative effect on the phenotype or development of the mouse compared to controls, because mixing of the alleles dilutes the potentially harmful nDNA-mtDNA effects (Chinnery *et al.*, 2014). These results suggest that the mitochondria alone are able to accommodate dramatic nuclear genetic changes, so long as the genetic code is available to support their function, and that mtDNA CN regulation may be an internal mitochondria and cellular response, rather than nuclear-mitochondrial genetic regulation.

Martin and colleagues used a new technique known as ChromEMT, which identified long range interactions between genes in nDNA in complex diseases (Martin *et al.*, 2015). The technique combines electron microscopy tomography with a labelling method that selectively enhances the contrast of DNA, which showed that nDNA has highly disordered nucleosomes which are packaged into 5-24nm diameters, in chromatic concentration volumes between 12% to 52% (depending on cell cycle stage), and with different structural and particle conformations. During interphase, chromatin is more curvilinear but collapses into compact, and interacting, loops in mitotic stages (Ou *et al.*, 2017). This provides evidence of the diverse array of different combinations of DNA sequence interactions, and how they are uniquely dependent on cell cycle phase in each individual. This, combined with chromatin concentration volumes, and individual genomic sequences, adds complexity to

understanding how nDNA controls cellular activities. It is possible to hypothesise that there may be multiple loci interacting during different cell cycle stages which are contributing to mitochondrial biogenesis regulation, and are not being identified through GWAS.

Reinecke and colleagues review the number of mitochondrial-related diseases that develop as a result of different expression profiles of OXPHOS genes under different cellular circumstances, which were dependent on OXPHOS, calcium signalling, and ROS (Reinecke *et al.*, 2009). mtDNA CN may be indicative under different cellular circumstances, and this may explain why our GWAS results don't replicate.

A GWAS of nearly 28,000 lincRNAs in 1829 samples also identified lincRNAs to be highly associated with genetic and expression control, as they originate from enhancers, and regulate transcription of their associated SNPs (Hon *et al.*, 2017). Additionally, approximately 88% of tagged SNPs lie in the intergenic or intronic regions and are likely to influence gene regulation (Edwards *et al.*, 2013). SNP functional consequences have been reviewed (Cooper and Shendure, 2011).

mtDNA may be marker of mitochondrial stress (Malik *et al.*, 2016), however it may not be vital for cellular function. Dean and colleagues showed that the mitochondria in parasites *Trypanosoma brucei,* which contained mutations in nDNA which code for ATP synthase, have the ability to maintain mitochondrial homeostasis and prevent cell death by reversing the ATP synthase pump to pump protons into the matrix to maintain the mitochondrial membrane potential (Δψm) (Dean *et al.*, 2013). This suggests Δψm is more important for mitochondrial cell survival than mtDNA to code for the machinery needed to maintain the Δψm.

4.5.5 Imputation Servers

This is the first study to investigate differences in the methodology between established imputation servers, and the quality of imputation reported, having imputed UKBS genotyping data using both the Sanger and Michigan Imputation Servers, and to both the HRC r1.0 and HRC r1.1 reference panels.

Very little difference was reported in the MAFs, imputation quality, or the number of variants which passed QC when comparing reference panel-imputed datasets (Figure 4.7 and Table 4.1). This suggests that there is little difference between reference panels HRC r1.0 and HRC r1.1. Differences in imputation scores were

reported between imputation servers. The Sanger Imputation Server reported a minimum imputation score of 0.4 across all variants, unless the MAF <0.01, whereas the Michigan Imputation Server reported a minimum imputation score of 0.1 across all variants, unless the MAF <0.01. This suggests there are differences in the mathematics behind the imputation methods, which would be expected because the Sanger Imputation Server uses PBWT methods, and Michigan Imputation Server uses Minimac3 methods. Figure 4.6 highlighted the differences in imputation scores by variants between the servers, reporting no correlation between their methods. However, this may be due to imputation methods imputing rarer variants better than the other. As Figure 4.7A shows, the Michigan Imputation Server had more variants removed based on imputation score alone, suggesting variants with a lower MAF (but not less than 0.01) were imputed better by the Sanger Imputation Server than by the Michigan Imputation Server. There is very little difference in the number of variants removed between server methods when removing variants under standard GWAS QC conditions (MAF <0.01, INFO <0.80) (Figure 4.7B).

Having analysed data from both imputation servers, manipulating Sanger-imputed data was much more user friendly than Michigan-imputed data, however both methods still produced the same set of results.

4.5.6 Study Limitations

Following the work in this study, many limitations have been identified which have contributed to the lack of replication in this work, but has also highlighted factors which other researchers should implement to ensure accurate GWAS reporting, especially in meta-analysis work.

4.5.6.1 Power

Our initial meta-analysis study was powered to find a single locus which controlled mtDNA CN with a MAF >0.10 and an effect size (β_G) of >4.8, up to a MAF >0.40 with a β_G of >3.0, assuming population means of 200 copies/cell, (±30 copies s.d). However this study, conducted in over 8000 individuals, indicates multiple loci are associated with changes, but with no replication. This suggests this study may be underpowered. Using the top list of genes identified to be associated with mtDNA CN in this study (Table 4.9), new power calculations were conducted using QUANTO to identify the real power of this study. Gene only, additive inheritance models in independent individuals were calculated.

The average beta (effect size or β_G) of the 13 genes reported in Table 4.9 is 0.5152 (maximum 2.316 for rs941898, minimum 0.0850 for rs8066582), and the average MAF is 0.1258 (maximum 0.4631 for rs8066582, minimum 0.01211 for rs140416442). Actual population mtDNA CN mean was 178 copies/cell (± 58 copies s.d). Assuming no change in our recorded population mean and standard deviation, in a worst case scenario, whereby β_G was 0.085 and MAF 0.012, >1,000,000 individuals would be needed to detect an effect at 5.00E-08 with 80% power. In a best case scenario, whereby the β_G was 2.316 and MAF 0.4631, >50,000 individuals would be needed to detect an effect at 5.00E-08 with 80% power. Using the current average beta and MAF values, this study has no power to detect an effect at 5.00E-08.

As these calculation suggest, and assuming there is no change in the population mean and standard deviation, to identify real nuclear genetic associations to mtDNA CN changes, much larger sample numbers are going to be needed. This is a result of the smaller population mean (178 copies/cell compared to the assumed 200 copies/cell prior to this study – section 2.11), and the large standard deviation recorded in mtDNA CN distribution (±58 copies compared to the predicted ±30 copies used to calculate power prior to this study – section 2.11). To reduce the deviation from the population mean, a single qPCR user would need to generate all mtDNA CN values to minimise technical variability. Kitchen and colleagues also show that the variance reported in qPCR measurements cannot be improved by using more technical replicates, and more biological replicates are needed to reduce the variance (Kitchen *et al.*, 2010). This supports my previous work in section 2.9.1, which shows the number of replicates used in qPCR does not improve the precision of the mtDNA CN measurement. Furthermore, a more accurate and repeatable qPCR assay needs to be designed to reduce mtDNA CN variation.

4.5.6.2 Technical

Several technically limiting factors were identified during this study with introduced limitations in our work.

4.5.6.2.1 Quantitative PCR

The first, and main technical limitation of this study were the different qPCR methods used to determine mtDNA CN in each cohort. UKBS samples used specifically targeted fluorescent TaqMan™ probes to quantify genes, however ALSPAC samples were quantified using SYBR® Green, which is less specific and binds to any double stranded DNA (Applied Biosystems, 2002). As Figure 4.16B suggests, Bristol's qPCR reaction may be plateauing out as a result of using up reagents too quickly, which may indicate that the qPCR assay needs to be optimised further. Additionally, primer designs need to ensure that the last five nucleotides at the 3´ end do not contain more than two G and/or C bases (Applied Biosystems, 2002), however the last five nucleotides of Bristol's mitochondrial primers 3' end contains four G and C bases (CCAGC). ALSPAC samples mitochondrial reference gene is also located in the D binding domain of mtDNA, which could potentially be affected by quadruplex stranding (J Nicholls and Minczuk, 2014). These technical limitations may explain why ALSPAC mtDNA CN values were four times lower than UKBS values, and although we tried to accommodate as much as possible to these variations, future studies need to conduct mtDNA CN measurements using the same qPCR assay for consistency (Kitchen *et al.*, 2010).

4.5.6.2.2 DNA Handling

It is well established that DNA handling contributes to DNA quality, which in turn affects the ability of qPCR to determine mtDNA CN. Freeze-thaw effects on blood samples are reported to decrease DNA integrity by 25% prior to DNA extraction (Ross *et al.*, 1990). This is because freezing is thought to mechanically break DNA, and ions, solutes, and metals present in water can create concentrated microenvironments during freezing which create localised pockets of pH changes which can also promote oxidation, and encourage DNA degradation. Adding cyroprotecting reagents to the DNA (like glycerol) when freezing maintains the DNA quality, even when being subjected to up to 16 freeze-thaw cycles (Schaudien *et al.*, 2007; Röder *et al.*, 2010), and the addition of Tris is thought to prevent pH-related damage (Graf *et al.*, 1984; Brunstein J., 2015).

DNA larger than 100kb is thought to be most sensitive to the freeze thaw effect (Shao *et al.*, 2012; Brunstein J., 2015), which suggests nDNA may be more susceptible to damage than mtDNA, and normalising mtDNA reference genes to *B2M* may be additionally compromised.

It is uncertain how much UKBS, and ALSPAC samples underwent freeze-thaw processes. Although we did the best possible to mitigate the effects in this study, a third cohort is needed to be measured at Newcastle to perform another GWAS to both increase the study power, and also account for the differences reported from qPCR assay effects.

4.5.6.2.3 Blood Count Data Determination

Some studies do not see a significant change in their mtDNA CN analysis when adjusting for blood cell count proportions, however this may be because their blood count measurements are estimates (Ameer *et al.*, 2016). Adjusting for cellular blood counts was an important factor in the analysis for this study, because mtDNA CN differences have been reported to differ by blood cell component both in this study (section 3.3.5), and in previous literature (Pyle *et al.*, 2010). Introducing blood count adjustments into the ALSPAC analysis introduced a lot of noise to the GWAS results (Figure 4.17B and F), however there was less noise in GWAS results using whole blood cell extracted samples (Figure 4.17D). Adjusting for cell counts also removed the association to the *PSMD3* loci on chromosome 17, which is associated with neutrophil count, in the all-ALSPAC, basic adjusted analysis. This suggests blood cell adjustments may only be necessary when using whole blood cell extracted DNA or when investigating mtDNA CN changes in studies. This has also been advised in other genetic association studies (Diroma *et al.*, 2014; Jaffe and Irizarry, 2014). Through a sex-specific linkage analysis, Lopez and colleagues also demonstrated that separate quantitative trait loci are involved in the control of peripheral blood mtDNA content in women and men, however this was only in 386 individuals (López *et al.*, 2012).

Furthermore, ALSPAC cell count data was generated using estimated values, and was not a direct measure of cell counts, like in UKBS samples. Data were only available for neutrophil and lymphocyte counts, and only in a few hundred ALSPAC samples (Table 4.6), which was not representative of the entire cohort. GWAS analysis was only adjusted to neutrophil and lymphocytes, and excluded eosinophil,

monocyte, and basophil counts. I previously reported significant differences in mtDNA CN in eosinophils (section 3.3.5 - Figure 3.14 and Table 3.4), which means they have not been accounted for in our study. What's more is regression coefficients between ALSPAC and UKBS blood cell components show UKBS blood cell proportions are more similar to ALSPAC blood coefficients, thereby rendering the fully adjusted analysis because cell proportions were incomparably adjusted (Table 4.10).

Future work would need to ensure full blood count data was available for all samples to enable correct adjustments.

ALSPAC

	Lymph	*Mono*	*Eos*	*Neu*
Lymph	1.00	0.05	0.38	-0.97
Mono	0.05	1.00	0.03	-0.17
Eos	0.38	0.03	1.00	-0.54
Neu	-0.97	-0.17	-0.54	1.00

UKBS

	LA	*MA*	*EA*	*NA*
LA	1.00	0.24	0.15	0.20
MA	0.24	1.00	0.16	0.19
EA	0.15	0.16	1.00	0.06
NA	0.20	0.19	0.06	1.00

	LP	MP	EP	NP
LP	1.00	0.07	0.05	-0.93
MP	0.07	1.00	0.12	-0.36
EP	0.05	0.12	1.00	-0.31
NP	-0.93	-0.36	-0.31	1.00

Table 4.10: Regression coefficients between blood cell components in ALSPAC and UKBS blood count data. LA: absolute lymphocyte count. MA: absolute monocyte count. EA: absolute eosinophil count. NA: absolute neutrophil count. LP: lymphocyte proportion. MP: monocyte proportion. EP: eosinophil proportion. NP: neutrophil proportion. Boxes indicate similar regression coefficients between lymphocytes (blue), and neutrophils (red).

4.5.6.3 Genetic

Genetic limitations may also be present in this study because UKBS and ALSPAC samples were genotyped on different Illumina arrays. Efficiencies of Illumina genotyping platforms may vary which will give rise to differences in genotyping quality, and may mean SNPs were called for differently (Guo *et al.*, 2013). We tried to mitigate these errors by imputing genotypes, and even though this increased genetic power, imputation does come with limitations because SNPs are imputed in linkage disequilibrium with one another, and may report SNPs which are not directly causal (Zheng *et al.*, 2011; Edwards *et al.*, 2013). What's more is multiple continuous traits are being adjusted for in our study, which could be pulling results in contradictory directions. It may have been more accurate to perform a dimension reduction style GWAS method, whereby multiple continuous traits are linearly combined to generate one value, which is then tested against the genotyping data (Wu and Pankow, 2015).

Furthermore, to account for population heterogeneity, ideally it would have been more accurate to use raw genetic data instead of summary data, whereby both populations could have undergone multidimensional testing to adjust for any population heterogeneity.

Despite these limitations, we accommodated to technical and genetic limitations as much as possible throughout our work. Limiting our datasets to only include samples from the same DNA source, and imputed our genetic data to the same reference panel, and filtered data to the same control measures. mtDNA CN assay validation tests were also performed to validate phenotype similarities, and blood cell components analysis enabled us to perform the most comparable GWAS possible with the data available.

In conclusion, this work has identified the complexities behind correctly conducting GWAS, and has explored factors which need to be accommodated for when performing meta-analysis between datasets from heterogeneous backgrounds. This work has also reiterated the complex nature of mtDNA CN regulation, and suggests there is no single locus governing its maintenance like Cai and colleagues suggest (Cai *et al.*, 2015). More studies are needed to expand on our work, and to increase the statistical power behind results reported. Furthermore, functional work would need to be conducted in cell lines and animal models to monitor the real effect nuclear genes have on mtDNA CN.

Chapter 5 Mitochondrial Genetic Factors Modulating Mitochondrial DNA Copy Number

5.1 Introduction

5.1.1 Mitochondrial DNA and Mitochondrial DNA Copy Number

The 16.6 kbp circular, double-stranded mitochondrial DNA (mtDNA) molecule encodes 13 out of approximately 85 subunits needed for the OXPHOS structural system (Anderson *et al.*, 1981; Wallace, 1999; Smeitink *et al.*, 2001). mtDNA was originally larger than it is now, and much of the original genome is believed to have translocated to the nuclear genome (nDNA) due to intracellular pressures (Brown *et al.*, 1979; Lane and Martin, 2010b; Gray, 2012; Johnston and Williams, 2016). mtDNA is a polyploidal genome, and is present in multiple copies within cells, however this is dependent on mtDNA quality and cellular metabolic demand (Lee and Wei, 2000; Liu *et al.*, 2003; Gilkerson, 2009). As a result, mtDNA CN varies considerably between cells (Michaels *et al.*, 1982; Robin and Wong, 1988; Kelly *et al.*, 2012).

5.1.2 OXPHOS and Mitochondrial DNA Copy Number

The OXPHOS system is made up of five complexes, and of the 13 OXPHOS proteins still encoded by mtDNA, seven code for complex I subunits, one codes for a subunit in complex III, three code for complex IV subunits, and two code for complex V subunits (Figure 5.1). The remaining OXPHOS subunits are coded for by nDNA. Lane and Martin (2010) believe that mtDNA has evolved to facilitate higher energy output for complex and multicellular organisms (Lane and Martin, 2010a), however uncertainty remains as to why mtDNA contains the specific genes it does. The “null hypothesis” is that gene loss has been uniform and random (Johnston and Williams, 2016). However, two other theories have been proposed that mtDNA is selected on qualitative measures (Daley and Whelan, 2005); whereby one hypothesises highly hydrophobic proteins are retained in mtDNA to bypass difficult transport across membranes in to the mitochondria, and the other, known as “colocalisation for redox regulation”, hypothesises that critical genes which are vital for localised control of energetic machinery are retained to avoid affecting other organelles. Further details are described here (Johnston and Williams, 2016).

mtDNA and OXPHOS have a close relationship because the cell’s energy demands are reliant on OXPHOS to generate ATP, and OXPHOS requires mtDNA replication and transcription to synthesise respective complexes. Dickinson and colleagues report a direct correlation between OXPHOS activity and mtDNA CN changes

(Dickinson *et al.*, 2013). LHON is a primary mitochondrial disease which usually develops as a result of having one of three common point mutations (*G3460A*, *G11778A* and *T14484C*) located in *MT-ND1*, *MT-ND4,* and *MT-ND6* of mtDNA respectively (Hudson *et al.*, 2007b). Not all carriers develop LHON, which is known as incomplete penetrance (Wallace *et al.*, 1988), and increased mtDNA CN is believed to be the protective mechanism (Giordano *et al.*, 2014). This suggests OXPHOS activity and mtDNA CN are regulated independently, but mtDNA CN is important because it is a mediator to disease development.

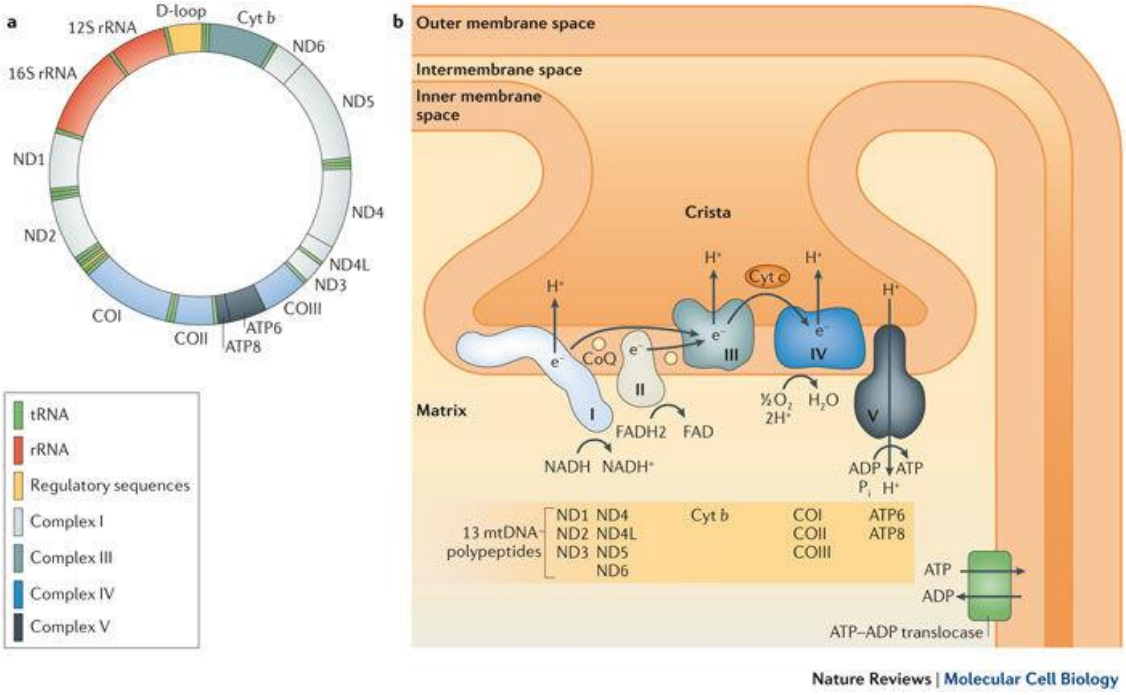


Figure 5.1: Illustration of mitochondrial DNA and the gene coding regions corresponding to their respective OXPHOS complex. Image taken from (Mishra and Chan, 2014).

5.1.3 Mitochondrial Haplogroups, Disease and Mitochondrial DNA Copy Number

Homo sapiens originated from Africa about 150,000 to 200,000 years ago (Cann et al., 1987; Wallace et al., 1999; Stewart and Chinnery, 2015). As a result of population migration, and exposure to environmental and diet changes, mtDNA accumulated mutations which are now seen in high frequencies between populations, and are polymorphic. These polymorphisms are associated with specific haplogroups, but all stem from the same ancestral tree. The most variation is reported in African haplogroups (Johnson et al., 1983), however mtDNA sequences between unrelated individuals in human populations can differ by about 50 nucleotides overall (Wallace *et al.*, 1999).

European haplogroups were defined by phylogenetic analysis which revealed 75% European haplogroups did not contain the *DdeI* site at nucleotide position 10394 (Torroni *et al.*, 1994; Torroni *et al.*, 1996). Haplogroups lacking the *DdeI* site are haplogroups H, T, U, V, W, and X, whereas haplogroups I, J, and K retained the *DdeI* site (Wallace *et al.*, 1999). Haplogroup R is defined root of European haplogroups, however haplogroup H is the most common haplogroup in Europe, and is comprised of approximately 90 different sub-haplogroups (Stewart and Chinnery, 2015).

Studies have reported specific mitochondrial haplogroups to have different OXPHOS capacities (Kofler *et al.*, 2009b; Gomez-Duran *et al.*, 2010; Hulgan *et al.*, 2011), which can affect muscle biogenesis (Nagao et al., 1998). Additionally, Northern European haplogroups are more tolerant to colder climates, with Southern European haplogroups appearing more tolerant to hotter climates (Ruiz-Pesini et al., 2004), demonstrating the different coupling effects of OXPHOS as a result of mtDNA coding.

Several studies have linked haplogroup to disease susceptibility (Pyle *et al.*, 2005; Hudson *et al.*, 2007b; Pello *et al.*, 2008; Kofler *et al.*, 2009a; Chinnery *et al.*, 2010; Gómez-Durán *et al.*, 2012; Soto-Hermida *et al.*, 2014), as well as protecting against disease (Ghezzi *et al.*, 2005; Hendrickson *et al.*, 2008; Santoro *et al.*, 2010; Shen *et al.*, 2011; Liou *et al.*, 2012; Hudson *et al.*, 2013a; Ridge *et al.*, 2014; Tranah *et al.*, 2014). Despite many metabolic-related diseases also being considered age-related diseases, the longevity association reported to mitochondrial haplogroup J (Rose et al., 2001; Dato et al., 2004), is not replicated across all populations (Ross et al., 2001; Niemi et al., 2003). Mueller's group used HEK cybrids to identify that haplogroup H and T had no OXPHOS or TCA enzyme differences, however haplogroup T cybrids had a significantly higher mtDNA CN (Mueller et al., 2012). This suggests that mtDNA variants may impact mtDNA CN regulation, but not OXPHOS efficiency.

5.1.4 Mitochondrial Haplogroups and Mitochondrial DNA Copy Number

Of the studies associating mitochondrial haplogroup to disease, some have identified relationships between mtDNA CN and disease development (Wang et al., 2012; Giordano et al., 2014). However, very few have reported the relationship between mtDNA variants and mtDNA CN changes in relation to the disease onset. For example, sepsis has been associated with haplogroups JT (Lorente et al., 2013), and

mtDNA CN has been reported to be lower in sepsis patients (Pyle et al., 2010). Also, reduced Parkinson's disease (PD) risk is associated with haplogroups UKJT (Pyle et al., 2005), and reduced mtDNA CN is a biomarker of PD (Pyle et al., 2015a; Pyle et al., 2015b). What's more, Ridge and colleagues reported haplogroups U5A1 and T2 to have significantly higher mtDNA CN in an aged cohort (Ridge et al., 2014). However, there are no studies to use as a reference which report mitochondrial haplogroup association to mtDNA CN changes in large European populations, or in healthy populations. To address this, and following on from previous GWAS work on nDNA to identify nuclear variants associated with mtDNA CN changes, I performed a GWAS using mitochondrial variants to identify mtDNA variants associated with mtDNA CN changes in two large, healthy, European cohorts.

5.2 Aim

Similar to my previous work (0), and based on previously published studies in disease (described above), I propose that inherited mtDNA variation is likely to modulate mtDNA CN.

The aim of this chapter is to perform two GWAS using mitochondrial SNPs (mtSNPs) in two independent control cohorts, finally meta-analysing the data in a combined analysis. Additionally, a third independent cohort will be used as replication of these results, and will also investigate qPCR assay method effects on genomic results.

5.3 Materials and Methods

5.3.1 Cohorts

5.3.1.1 Cohort One - UKBS

Full cohort details and methods are available in Chapter 2. For the work in this chapter, 3091 UKBS samples (50% male, 50% female) were used as a discovery cohort. UKBS DNA samples were extracted from white blood cell pellets using phenol-chloroform methods, and FBC data was provided for 2617 samples. Relative mtDNA CN was determined using an established triplex TaqMan™ qPCR assay, quantifying mitochondrial genes *MT-ND1* and *MT-ND4*, to nuclear housekeeping gene *B2M* (Grady *et al.*, 2014a; Pyle *et al.*, 2015b; Rygiel *et al.*, 2015). 2735 samples had 132 mtDNA SNPs (mtSNPs) (>99%) successfully genotyped on the on the Illumina 1.2M Duo array (Illumina, San Diego, California, USA), of which 2682 (50% males, 50% females) had both genotyping and mtDNA CN data.

5.3.1.2 Cohort Two – In house 'Cohort Two'

Full cohort details and methods are available in Chapter 2. For the work in this chapter, 2000 'cohort two' samples were used as an independent, in house replication cohort. DNA samples were extracted from blood (94.45%) and buccal (5.55%) samples from 1000 mother-child pairs. FBC data, DNA extraction method, age, and gender of samples was not available at the time of analysis. Relative mtDNA CN was determined using the same qPCR assay as cohort one. 1627 samples had their entire mtDNA sequenced by Source BioScience™ in Nottingham, UK on the Illumina HiSeq 2000™ array (Illumina, San Diego, California, USA). 1605 samples (80.25%) had both mtDNA sequencing and mtDNA CN data.

5.3.1.3 Cohort Three - ALSPAC

Full cohort details and methods are available in Chapter 2. For the work in this chapter, data for 5461 ALSPAC maternal samples (100% female) was available through collaboration with Dr Santi Rodriguez's group at University of Bristol, and was used as an independent replication cohort. DNA samples were previously extracted from whole blood (37.6%) or WBC (62.4%) using a phenol-chloroform method (Fraser *et al.*, 2013), and relative mtDNA CN was determined using a singleplex SYBR® Green assay which quantified a region of the mitochondrial D-loop to a region of nuclear housekeeping gene *B2M.* Estimated blood cell count data was

also available for 482 samples. mtSNP genotyping data was assigned to mothers using their respective child's data (as mtDNA is inherited maternally (Hutchison *et al.*, 1974; Giles *et al.*, 1980)) which had been genotyped directly on the Illumina HumanHap550-Quad array (Illumina, San Diego, California, USA). 3405 mothers had mtSNP genotyping and mtDNA CN data. All analysis for cohort three was conducted at University of Bristol, and summary data was provided to Newcastle University to perform respective meta-analysis'.

5.3.2 Mitochondrial Haplogroup Assignment

5.3.2.1 Cohort One - UKBS

The hierarchical relationship amongst mtDNA variants is represented at www.phylotree.org. A schematic overview of the mitochondrial haplogroup phylotree is shown in Figure 5.3. To investigate the effect mitochondrial genetic variants have on mtDNA CN differences, UKBS individuals were assigned to their European mitochondrial haplogroup according to published criteria (Torroni *et al.*, 1996). The mtSNPs present on the Illumina 1.2M array (detailed in section 2.6.1) were first converted from Yoruban to rCRS format using MITOMAP (https://www.mitomap.org/MITOMAP/YorubanConversion), and then haplogroups were assigned using specific genotyped variants (Figure 5.2). For example haplogroup J is defined by variants *C295T* and *T489C*. If it was not possible to call a major-haplogroup defining SNP then clade specific sub-type variants were used to identify mtDNA haplogroups. All samples were assigned a haplogroup. To investigate mtDNA CN differences between major European haplogroups, non- European haplogroups (non-RHVJTUKWXI haplogroups) were removed (n=45) as a final QC metric to avoid the effects of phylogenetic heterogeneity.

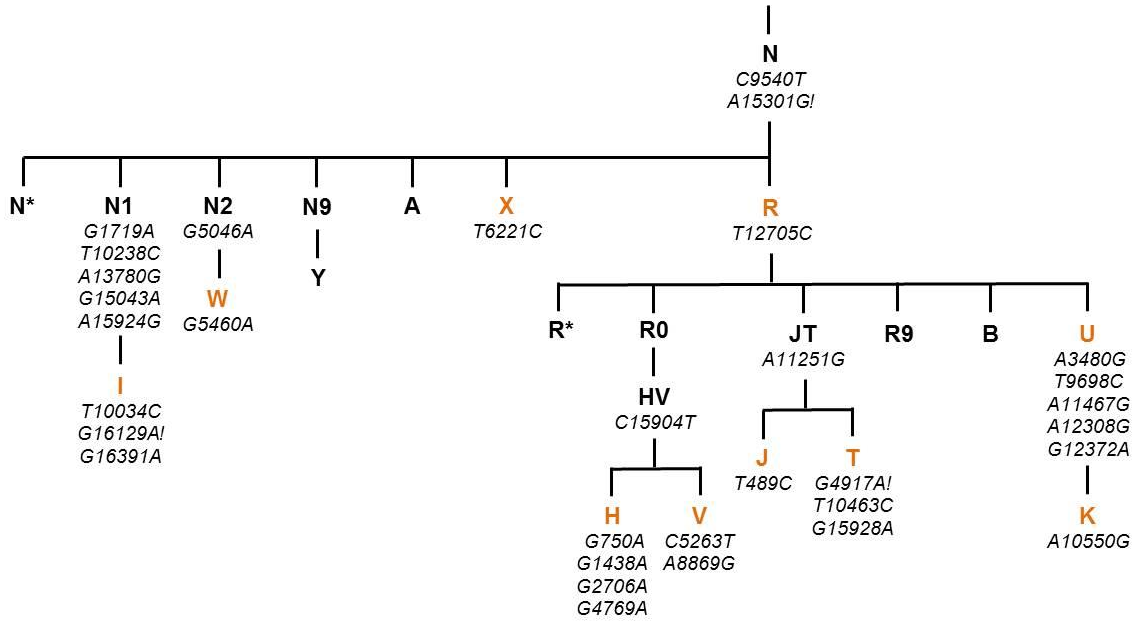


Figure 5.2: Non-rCRS SNPs on the Illumina 1.2M Dual genotype array used to call for specific mitochondrial haplogroups in UKBS samples. For example, haplogroup J genotyping would ideally need to call for C9540T, T12705C, A11251G and T489C. Illumina SNPs have been converted from Yoruba to rCRS positions using https://www.snpedia.com/index.php/MtDNA_Position_Conversions.

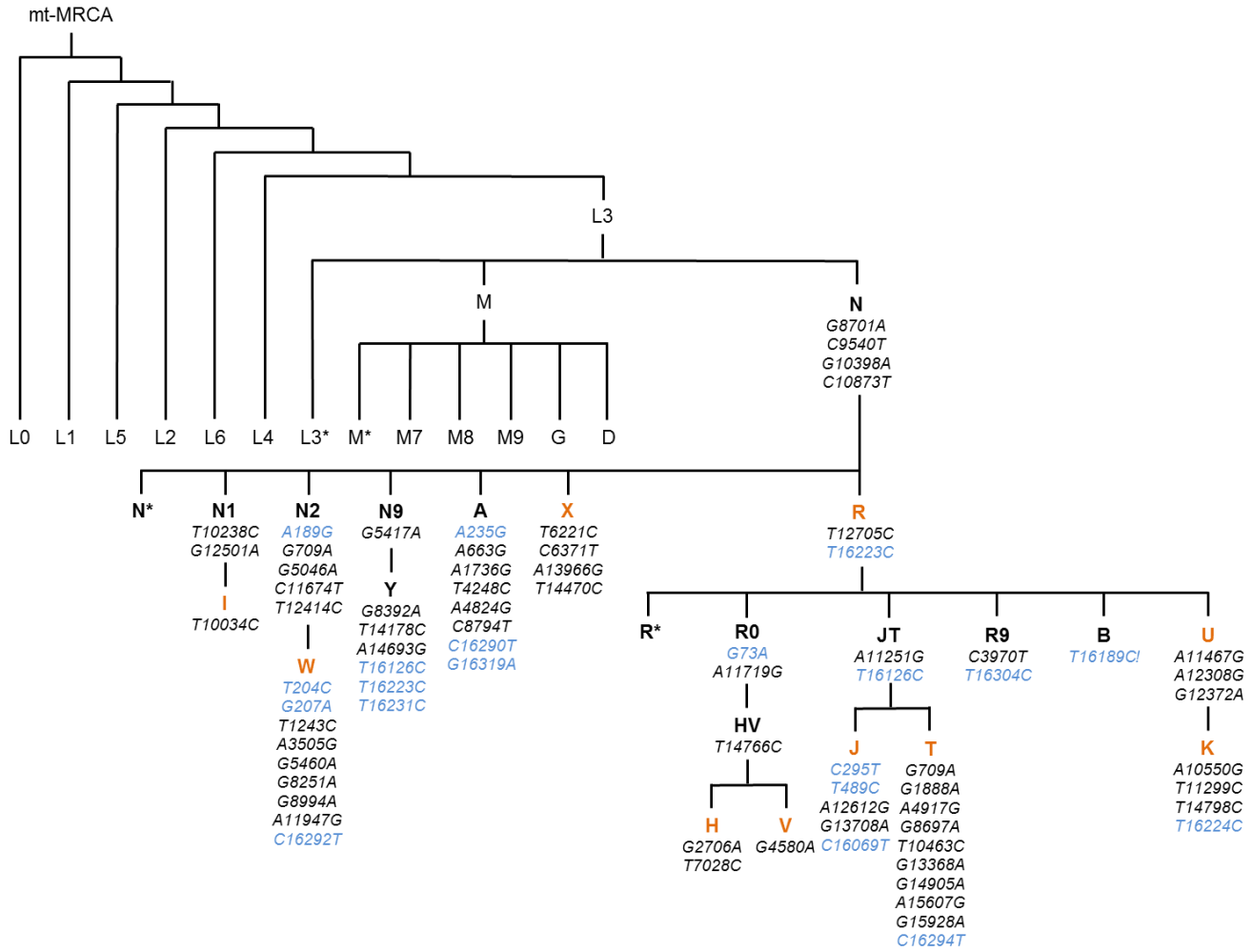


Figure 5.3: Schematic representation of the human mtDNA phylogenetic tree. Each branch details the evolution of mtDNA which stemmed from mitochondrial eve (mt-MRCA). Stable nucleotide polymorphisms in mtDNA give rise to specific mitochondrial haplogroups. The European haplogroup clade is detailed below. Haplogroups highlighted in orange are common European haplogroups. Canonical genomic variants below each branch usually define the haplogroup (Torroni et al., 1996). Variants in blue sit in the control region of mtDNA. Image is adapted from the revised Cambridge Reference Sequence (rCRS) (Build 17 (18 Feb 2016)) - http://www.phylotree.org/index.htm (van Oven and Kayser, 2009).

5.3.2.2 Cohort Two – In house 'Cohort Two'

Cohort two samples were effectively sequenced for mtDNA variation by Source BioScience (Nottingham, UK) (detailed in section 2.6.2). Mitochondrial haplogroups were also identified using HaploGrep2 (http://haplogrep.uibk.ac.at/) (van Oven and Kayser, 2009; Kloss-Brandstätter et al., 2011; van Oven, 2015; Weissensteiner et al., 2016), and were confirmed using MitoMaster (https://www.mitomap.org/foswiki/bin/view/MITOMASTER/WebHome). 1627 samples (81.35%) had their mtDNA successfully sequenced and haplogroup defined. As part of a further QC step, mitochondrial haplogroups were manually assigned to identify the accuracy of HaploGrep2 and MitoMaster. 45 individuals were assigned the incorrect haplogroup, suggesting the software is approximately 98% accurate.

5.3.2.3 Cohort Three - ALSPAC

ALSPAC samples were genotyped on the Illumina HumanHap550 quad genotyping array by 23andMe (detailed in section 2.6.3), and mitochondrial haplogroups were assigned using HaploGrep2 (http://haplogrep.uibk.ac.at/) which automatically determines mitochondrial haplogroups using Phylotree (van Oven and Kayser, 2009; Kloss-Brandstätter *et al.*, 2011; van Oven, 2015; Weissensteiner *et al.*, 2016). Samples with a quality score >90% were used for analysis. All mtDNA analysis and haplogroup assignment was conducted by Dr Mesut Erzurumluoglu during his PhD with Dr Santi Rodriguez at University of Bristol.

5.3.3 GWAS Analysis

5.3.3.1 Cohort One - UKBS

As previously described in section 4.3.2.1, cryptic relatedness tests (IBD score >0), and multidimensional scaling analysis confirmed individuals were unrelated, and were from European descent (Burton *et al.*, 2007). Further filtering of mtDNA genotyped data removed individuals with missingness >3%, and variants with missing call rates >5%, and MAF <1%. UKBS were then stratified by gender, and QC filters applied to each group. 2520 people (1256 female, 1264 male) and 67 mtSNPs passed QC, with a genotyping rate 0.994170. Binary files were then filtered to remove non-European haplogroup (M, L, and non-N) individuals (n=45). 2469 individuals (1228 females, and 1241 males) had mtSNP genotyping and mtDNA CN data. 2145 (1073 females, and 1072 males) had additional blood count data.

Linear regression association tests were then performed, using logged, z-scored mtDNA CN as a continuous, quantitative phenotype. No additional filters were included in these tests. Absolute neutrophil and lymphocyte counts were included as covariates for cell-count adjusted analysis (Table 5.1). All analysis and filtering was performed using PLINK v1.90b3.44 64-bit (17 Nov 2016) (Purcell S *et al.*, 2007).

5.3.3.2 Cohort Two – In house 'Cohort Two'

Due to time constraints of the project, cohort two was not used to perform a GWAS, but was used to replicate a haplogroup effect instead of a genotype effect.

5.3.3.3 Cohort Three - ALSPAC

3405 mothers and 105 mitochondrial SNPs, with a MAF >0.01, passed QC checks detailed in section 2.6.3. Linear regression association tests were performed, using logged, z-scored mtDNA CN as a continuous, quantitative phenotype. No additional filters were included in these tests. Absolute neutrophil and lymphocyte counts were included as covariates for cell-count adjusted analysis.

5.3.4 Meta-GWAS Analysis

UKBS and ALSPAC GWAS summary data was used to perform a meta-analysis, however summary files were manually edited prior to analysis. To make analysis consistent between cohorts, each original mtSNP had a new ID assigned to it. New mtSNP IDs were "chromosome_alleleA_baseposition_alleleB", whereby allele A was A2, and allele B was A1 in UKBS PLINK output files. ALSPAC summary file alleles were the wrong way around and were therefore reordered to match UKBS alleles. Summary files were then filtered to only include mtSNPs that were present in both cohorts (Table 5.1). As variants were not imputed, an imputation score of one was assigned to each variant.

	Number of Individuals (N)			Mitochondrial SNPs		
Cohort Analysis	**Original**	**Phenotype + Genotype**	**Cell Count data**	**Original**	**Removed**	**Remaining**
All UKBS	3091	2469	2145	67	14	53
ALSPAC (WC)	3405	2117	264	105	52	53
UKBS FEMALES	1550	1228	1073	64	13	51
ALSPAC (WC)	3405	2117	264	105	54	51

Table 5.1: Number of individuals and mitochondrial SNPs used in meta-analysis between UKBS and ALSPAC cohorts. WC: white blood cell extracted samples only.

All UKBS samples (n=2469) were meta-analysed with white blood cell extracted ALSPAC (n=2117) samples, and UKBS females (n=1228) were meta-analysed with white blood cell extracted ALSPAC (n=2117) samples. Both meta-analysis was performed twice, with and without cell count adjusted summary data (Table 5.1). Meta-analysis was performed using META version 1.7 software (Marchini *et al.*, 2007; de Bakker *et al.*, 2008; Liu *et al.*, 2010) using the inverse-variance method based on a fixed-effects model.

5.3.5 Statistical Analysis

Linear regression GWAS tests were performed using PLINK v1.90b3.44 64-bit (17 Nov 2016) (Purcell S *et al.*, 2007). Independent T-tests (ITT) and Mann- Whitney U (MWU) tests were conducted using IBM SPSS Statistics v.22 (IBM Corporation, 2013), and Windows Excel 2013.

5.3.6 Quantifying Mitochondrial DNA Copy Number using an 'Alternative' Quantitative PCR Methodology

As discussed in section 5.4.6, an alternative qPCR assay was designed and an alternative region of mtDNA (*MT-RNR2*) was measured to quantify mtDNA CN. This assay was designed to validate the accuracy of the previously established *MT-ND1, MT-ND4,* and *B2M* assay detailed in section 2.8.1. Further justification for using this 'alternative' qPCR assay is detailed in section 5.4.6.

5.3.6.1 Cohort One - UKBS

Full methodology details are in section 2.8.1. All PCR and qPCR methods were exactly the same, however *MT-RNR2* was measured instead of *MT-ND1*. *MT-RNR2* DNA standard templates were generated and extracted as detailed in section 2.8.1.1, and *MT-RNR2* was quantified relative to nuclear housekeeping gene *B2M* and *MT-ND4* in a selection (n=108) of previously diluted UKBS samples (section 2.7.1) using the previously detailed triplex TaqMan™ qPCR assay (section 2.8.1.5) methods. The *MT-RNR2*-HEX fluorophore probe was designed to bind at bases 2980 to 3005. PCR and qPCR primer details are in Table 2.3 and Table 2.4 respectively. Each UKBS sample was measured in quadruplicate with a standard curve, control, and deletion calibre DNA. UKBS samples were measured across three separate qPCR runs across two, consecutive days. mtDNA CN was calculated for each sample by determining the relative proportion of *MT-RNR2* to *B2M* using Equation 2.2 in section 2.8.1.5. 105 samples (97.22%) had mtDNA CN values calculated.

5.3.6.2 Cohort Two – In house 'Cohort Two'

Full methodology details are in section 2.8.1. A selection (n=40; n=10 mothers and n=30 children) of cohort two samples had their mtDNA CN redetermined using the 'alternative' qPCR – detailed in section 5.3.6.1. Cohort two samples were also measured in quadruplicate, and samples were measured on one qPCR run on the same day they were plated out (i.e. they were not frozen).

5.3.6.3 Cohort Three – ALSPAC

ALSPAC samples were not available at Newcastle University and therefore were not measured using the alternative qPCR method.

5.4 Results

5.4.1 Mitochondrial DNA Copy Number Distributions in Study Cohorts

mtSNP genotyping data was available for UKBS and ALSPAC cohorts (as detailed in section 2.6), and mitochondrial haplogroups were assigned to samples in all cohorts (UKBS, ALSPAC, and 'Cohort Two'). Before analysis was conducted, frequency distributions of mtDNA CN were generated for all cohorts to investigate phenotypic differences (Figure 5.4).

As Figure 5.4 [A] – [G] show, mtDNA CN distributions were negatively skewed in all cohorts, however logging and z-scoring values removed this, resulting in normal distribution curves which were relative between cohorts (Figure 5.4 [a] – [g]). Therefore it was agreed that logged, z-scored mtDNA CN would be used for future analysis.

5.4.2 Cohort One - UKBS as a Discovery Study

5.4.2.1 Mitochondrial SNP Variants Associated with Mitochondrial DNA Copy Number

Acting as a discovery study, UKBS samples mitochondrial SNP genotyping data was cleaned, and linear association tests were performed on all UKBS samples (n=2520), UKBS females only (n=1256), and UKBS males only (n=1264). Analysis was performed with or without adjusting for blood cell counts. As Table 5.2 shows, two mtSNPs – *G5046A* and *G5460A* (both call for haplogroup W), were significantly associated with mtDNA CN changes across all analysis. mtSNP *T12705C* (which calls for haplogroup R) was significant in UKBS females but not males, and mtSNPs *G1719A* (which calls for haplogroup N1), *G12372A* (which calls for haplogroup U), *T6221C* (which calls for haplogroup X), and *G15043A* (which calls for haplogroup I), were significant in UKBS males but not females. *G1719A* and *G15043A* were not significant in males when cell count was adjusted for.

mtSNPs *G5046A* and *G5460A* were both used to call for mitochondrial haplogroup W in UKBS samples, and as Table 5.2 shows, they both have a negative effect on mtDNA CN. This supports previous results reported in Figure 5.5.

5.4.2.2 Mitochondrial DNA Copy Number Differences between Mitochondrial Haplogroups

UKBS samples had their mitochondrial haplogroup assigned using mtSNP data (detailed in section 5.3.2). To investigate whether mtDNA CN was different between mitochondrial haplogroups, UKBS samples (n=2520) were stratified by mitochondrial haplogroups, and bar graphs were generated showing mean mtDNA CN per haplogroup (Figure 5.5A). UKBS samples were then stratified by gender to identify if mtDNA CN differences by mitochondrial haplogroup were gender specific (Figure 5.5 B and C). As Figure 5.5 indicates, mitochondrial haplogroup W is significantly lower than other European mitochondrial haplogroups, and is not gender specific.

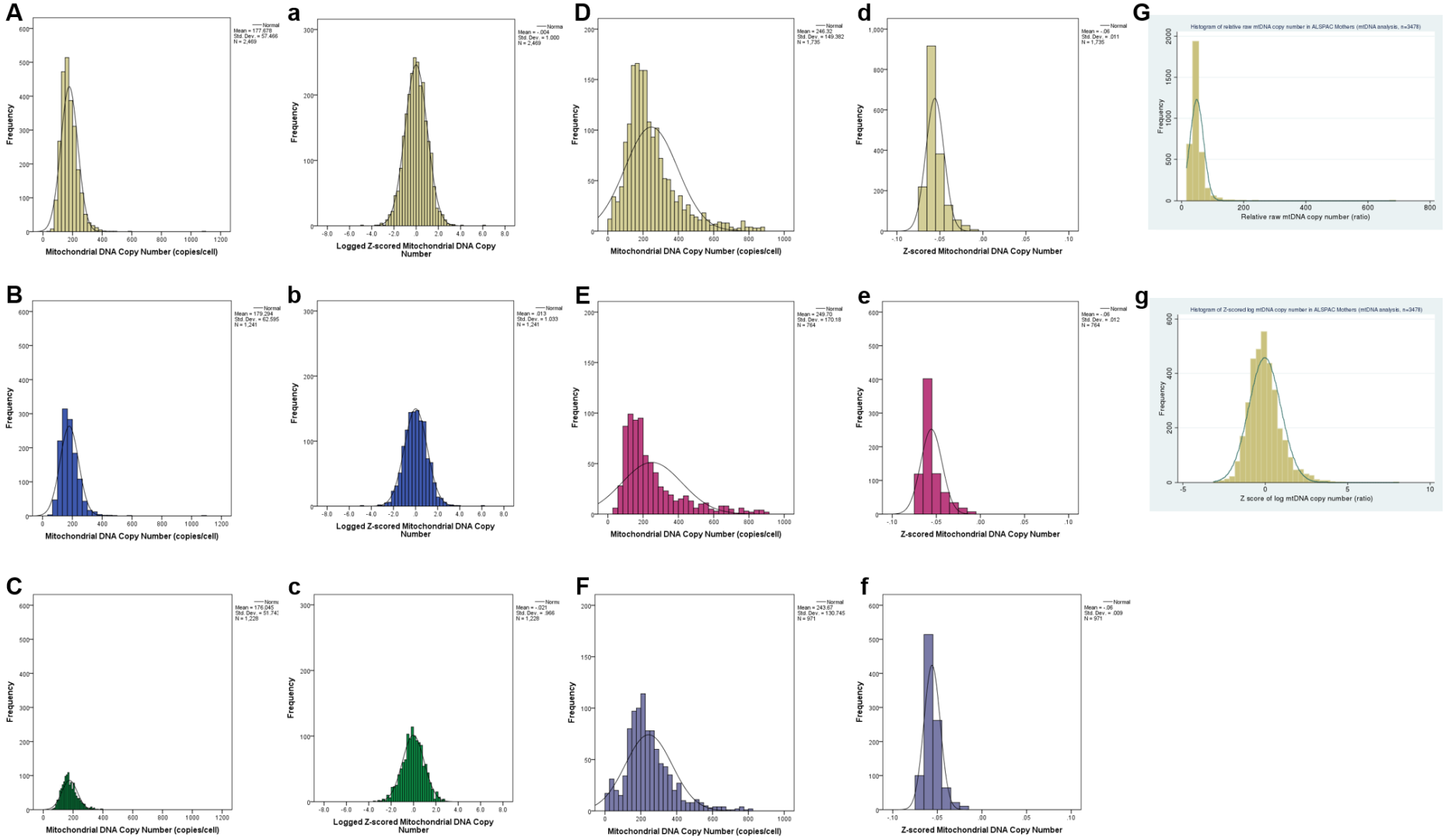
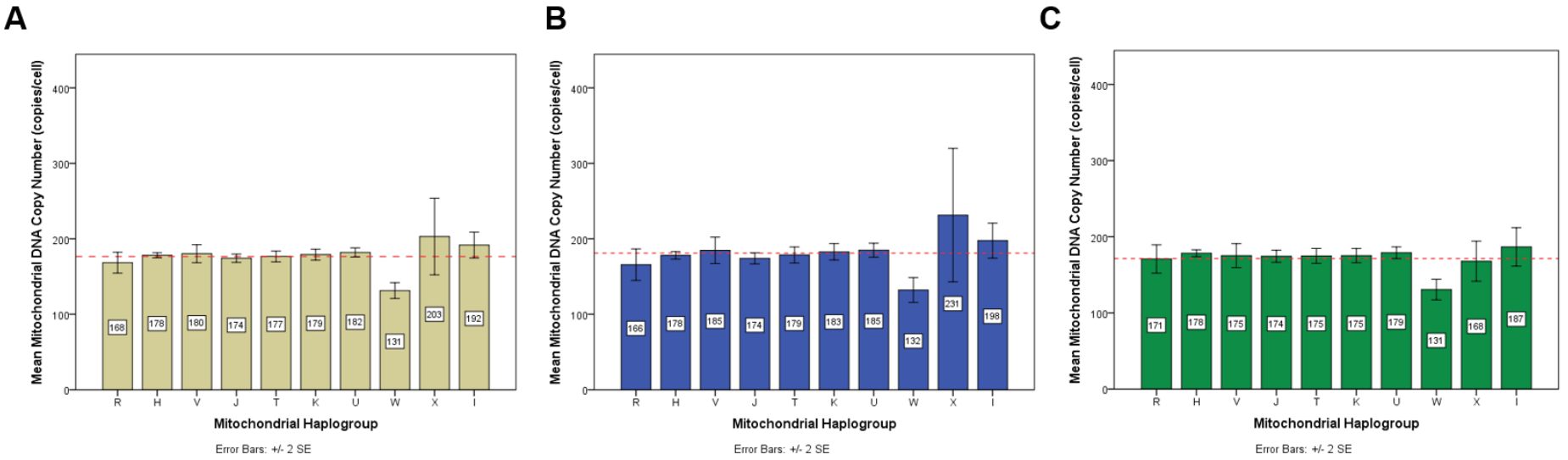


Figure 5.4: Frequency distribution plots of [A-G] mitochondrial DNA copy number (copies/cell) and [a-g] logged z-scored mitochondrial DNA copy number in UKBS, cohort two, and ALSPAC cohorts. [A, a] All UKBS (n=2520). [B, b] UKBS males only (n=1264). [C, c] UKBS females only (n=1256). [D, d] All cohort two (n=1735). [E, e] Cohort two mothers only (n=764). [F, f] Cohort two children only (n=971). [G, g] ALSPAC mothers (n=3405) – graphs provided by Anna Guyatt, Dr Santi Rodriguez's group, University of Bristol, UK.

		All (n=2469)							Females Only (n=1228)							Males Only (n=1241)							
		Not Cell Count Adjusted			Cell Count Adjusted				Not Cell Count Adjusted			Cell Count Adjusted				Not Cell Count Adjusted			Cell Count Adjusted				
SNP ID	rCRS BP	BETA	SE	P	BETA	SE	P	MAF	BETA	SE	P	BETA	SE	P	MAF	BETA	SE	P	BETA	SE	P	MAF	Associated Haplogroup
MitoG5047A	5046	-0.984	0.140	2.95E-12	-0.996	0.143	4.44E-12	0.021	-0.976	0.186	1.91E-07	-0.905	0.200	6.89E-06	0.022	-0.991	0.211	2.98E-06	-1.087	0.205	1.29E-07	0.019	L1b1, L3d1b, M63, M19, N2, B5a2a2a
MitoG5461A	5460	-0.408	0.096	2.12E-05	-0.476	0.100	2.05E-06	0.046	-0.367	0.132	5.49E-03	-0.360	0.143	1.17E-02	0.046	-0.448	0.139	1.27E-03	-0.586	0.140	3.22E-05	0.047	Homoplastic[1]
MitoT12706C	12705	-0.243	0.089	6.25E-03	-0.308	0.092	8.63E-04	0.054	-0.400	0.121	9.85E-04	-0.435	0.129	7.87E-04	0.055								L0g, R
MitoG1721A	1719	0.207	0.103	4.47E-02				0.040								0.316	0.145	2.94E-02				0.043	Homoplastic[2]
MitoG12373A	12372	0.096	0.048	4.66E-02	0.100	0.049	4.31E-02	0.229								0.141	0.070	4.35E-02	0.146	0.071	4.16E-02	0.231	M5a2a1a1, M7c1a1, M12, D4h1a1, N9a, A2ad1, H3au, U
MitoT13966C	13965				0.399	0.1993	4.54E-02	0.011															M7b1a2a1b, C1b13d, T2a
MitoT6222C	6221															0.4811	0.2219	3.03E-02	0.4937	0.2331	3.44E-02	0.017	L1c3, L3b, L3e1, M66a, C4a1a6, G3a2a, Q1c, O1, X, B4c2b, U7a4a1,
MitoG15044A	15043															0.4445	0.2082	3.30E-02				0.020	L2c2a, M, N1a1, J1c2a3, T2f1a, U6a7, U2c1a,

Table 5.2: Summary of top mitochondrial SNP hits (p< 0.05) for linear GWAS tests performed on UKBS samples for the association of mitochondrial variants effect on mitochondrial DNA copy number variability. Data was stratified by gender then tests were performed with or without cell count (absolute lymphocyte count and absolute neutrophil count) adjustments. rCRS BP= revised Cambridge Reference Sequence base position, SE= standard error, MAF= minor allele frequency. [1]Associated with 41 clades. [2] Associated with 33 clades.



D

	All UKBS (n=2520)					UKBS Males (n=1264)					UKBS Females (n=1256)				
Haplogroup	*N*	*Mean CN*	*SE*	*U95*	*L95*	*N*	*Mean CN*	*SE*	*U95*	*L95*	*N*	*Mean CN*	*SE*	*U95*	*L95*
All	2469	177.68	1.16	179.95	175.41	1241	179.29	1.78	182.78	175.81	1228	176.05	1.48	178.94	173.15
∟ R	51	168.35	6.88	182.17	154.53	24	165.60	10.41	187.14	144.06	27	170.79	9.29	189.88	151.70
∟ H	1104	178.21	1.69	181.52	174.90	547	178.11	2.51	183.05	173.17	557	178.30	2.26	182.73	173.87
∟ V	84	180.25	5.94	192.06	168.44	45	184.63	8.78	202.32	166.94	39	175.19	7.85	191.09	159.29
∟ J	304	174.22	2.70	179.55	168.90	167	174.09	3.71	181.42	166.77	137	174.38	3.96	182.22	166.55
∟ T	237	176.63	3.61	183.74	169.52	112	178.65	5.33	189.21	168.10	125	174.83	4.92	184.56	165.09
∟ U	356	181.94	3.00	187.85	176.03	181	184.84	4.60	193.93	175.76	175	178.93	3.83	186.50	171.36
∟ K	201	178.96	3.57	186.00	171.91	100	182.73	5.43	193.50	171.97	101	175.22	4.66	184.46	165.98
∟ W	51	131.35	5.21	141.83	120.88	24	132.17	8.22	149.17	115.17	27	130.62	6.75	144.50	116.74
∟ X	38	202.89	25.37	254.29	151.48	21	231.30	44.19	323.49	139.12	17	167.79	13.11	195.59	139.98
∟ I	43	191.71	8.57	209.01	174.40	20	197.62	11.64	221.99	173.25	23	186.57	12.57	212.64	160.50

Figure 5.5: Bar graphs showing the mean mitochondrial DNA copy number for each European mitochondrial haplogroup in UKBS samples. [A] All UKBS (n=2520). [B] UKBS males only (n=1264). [C] UKBS females only (n=1256). [D] Table summarising mitochondrial DNA copy number statistics for each respective mitochondrial haplogroup in each stratified UKBS group. CN: mitochondrial DNA copy number. SE: standard error. U95: upper 95% confidence interval. L95: lower 95% confidence interval. Red dotted lines indicate population means.

5.4.3 Cohort Three - ALSPAC as a Replication Study

5.4.3.1 Ensuring Mitochondrial Genotyping Data is Comparable at the SNP level

In collaboration with Dr Santi Rodriguez's group at University of Bristol, ALSPAC was used as a replication cohort. MAFs were generated for ALSPAC and UKBS datasets, using PLINK, to investigate population heterogeneity. mtSNP base positions were converted from Yoruban to rCRS format using MITOMAP (https://www.mitomap.org/MITOMAP/YorubanConversion), then mutual cohort mtSNPs MAF were compared using linear regression analysis (Figure 5.6).

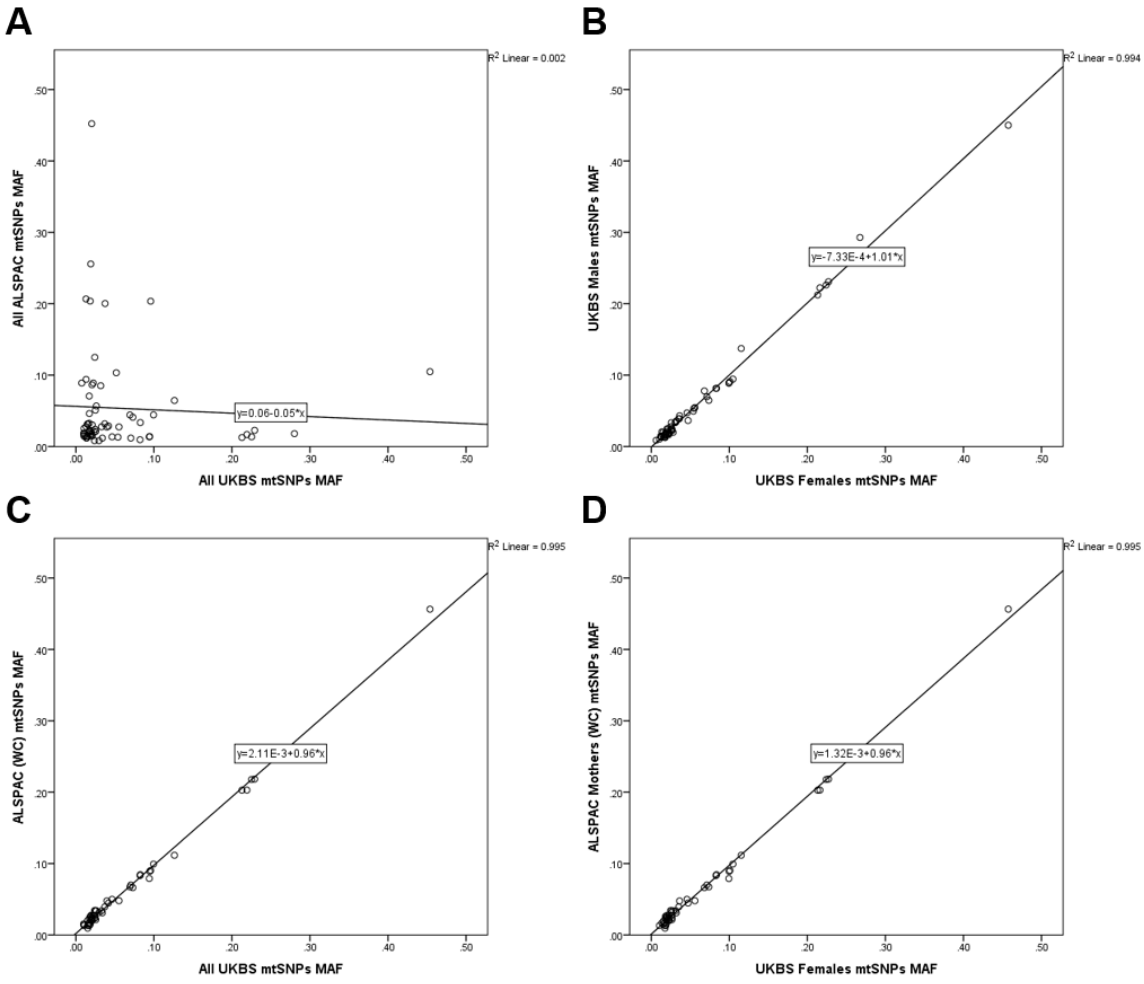


Figure 5.6: Scatter plots of the minor allele frequencies (MAF) of mitochondrial SNPs (mtSNPs) in ALSPAC (n=3405) and UKBS (n=2520) cohorts. [A] All UKBS versus all ALSPAC mtSNPs (n=67). [B] UKBS females versus UKBS males mtSNPs (n=62). [C] All UKBS versus white blood cell (WC) extracted ALSPAC samples mtSNPs (n=53). [D] UKBS females versus WC extracted ALSPAC samples mtSNPs (n=51). Lines indicate linear regression relationship.

As reported in Figure 5.6A, mtSNP MAFs between all ALSPAC (n=3405) and UKBS (n=2520) samples do not correlate (r^2= 0.002, β= -0.05), however mtSNPs in UKBS females (n=1256) and males (n=1264) positively correlate (r^2= 0.994, β= 0.997) (Figure 5.6B). ALSPAC samples were restricted to samples extracted from WBC only (n=2117), and mtSNP MAFs were compared to all UKBS samples (n=2520) (Figure 5.6C), or UKBS females (n=1256) only (Figure 5.6D). A positive correlation was reported between both analysis (r^2= 0.995, β= 0.997) and (r^2= 0.995, β= 0.998) respectively. This indicated that all UKBS samples, and UKBS females were suitable to meta-analyse with ALSPAC mothers extracted from WBC to identify mtSNPs associated with mtDNA CN.

5.4.3.2 Mitochondrial SNP Variants Associated with Mitochondrial DNA Copy Number

ALSPAC samples from University of Bristol were used as a replication study. Mitochondrial SNP genotyping data was cleaned, and linear association tests were performed on all ALSPAC maternal samples (n=3405), and then maternal samples extracted from whole blood (n=1284), and maternal samples extracted from white blood cell pellets (n=2117). Analysis was performed with or without adjusting for blood cell counts.

As Table 5.3 shows, 26 mtSNPs were significantly associated with mtDNA CN changes across different analysis. No mtSNP was consistent across all analysis, however mtSNPs *T489C* and *C16069T* (which both call for haplogroup J), and *T5999C* and *A6047G* (which both call for haplogroup U4), were consistently significant for analysis on whole blood cell extracted samples, without cell count adjustment. *A15218G* and *A13617G* (which both call for haplogroup U5) were consistently significant for analysis on whole blood cell extracted samples, with cell count adjustment. On the contrary, mtSNPs *G8269A* and *C242T* (which both call for haplogroup J1b) were consistently significant for analysis on white blood cell extracted samples, without cell count adjustment, whereas *G5046A* and *A11947G* (which both call for haplogroup W) were consistently significant for analysis on white blood cell extracted samples, with cell count adjustment.

5.4.3.3 Mitochondrial DNA Copy Number between Mitochondrial Haplogroups

GWAS analysis performed on ALSPAC samples produced results that were inconsistent overall (Table 5.3). However, results from white blood cell extracted samples were compared to UKBS samples, which were also extracted from white blood cells, making the cohorts comparable. Non-cell count adjusted analysis in ALSPAC samples had two significant results for mtSNPs *G8269A* and *C242T* (which both call for haplogroup J1b), which were negatively associated with mtDNA CN. This did not replicate UKBS results (Figure 5.5). However, cell count adjusted analysis in ALSPAC samples had significant results for mtSNPs *G5046A* and *A11947G* (which both call for haplogroup W), which were negatively associated with mtDNA CN. This replicated the results reported in UKBS samples (Figure 5.5).

			All							Whole Blood Extracts							White Blood Cell Extracts							
			Not Cell Count Adjusted (n=3401)			Cell Count Adjusted (n=482)				Not Cell Count Adjusted (n=1284)			Cell Count Adjusted (n=218)				Not Cell Count Adjusted (n=2117)			Cell Count Adjusted (n=264)				
SNP ID	ALSPAC BP	rCRS BP	BETA	SE	P	BETA	SE	P	MAF	BETA	SE	P	BETA	SE	P	MAF	BETA	SE	P	BETA	SE	P	MAF	Associated Haplogroups
snp2460C_new2587	8270	8269	-0.152	0.047	1.25E-03				0.030								-0.179	0.061	3.57E-03				0.034	M61, D4h2, N14, X2b3, H4a1a, J1b, U5a1a2b1
snp18251_new1669	242	242	-0.226	0.074	2.22E-03				0.012								-0.248	0.095	9.52E-03				0.013	J1b1a, U5b1b1g1a,
MitoT491C	491	489	-0.064	0.027	1.56E-02	-0.159	0.074	3.27E-02	0.105	-0.109	0.039	5.20E-03				0.094								M, J, U2e1f1,
snp18544_new1141	16070	16069	-0.063	0.026	1.67E-02	-0.170	0.073	1.97E-02	0.105	-0.089	0.038	2.03E-02				0.094								L0d2b, M30c1a, M91b, D5b3a1, N10b, HV4b, J
snp2172C_new2262	6000	5999	0.122	0.054	2.36E-02				0.022	0.236	0.077	2.36E-03				0.021								M91a, H3w, U4'9
snp7990_new2272	6048	6047	0.123	0.055	2.64E-02				0.021	0.226	0.079	4.15E-03				0.020								U4
snp28503_rs28359168	11948	11947				-0.413	0.147	5.07E-03	0.021	-0.183	0.092	4.72E-02				0.015				-0.444	0.217	4.14E-02	0.023	W, U6d3a
snp2706C_new959	15219	15218				0.293	0.105	5.58E-03	0.041				0.322	0.131	1.47E-02	0.037								M7a1a2, M10a1, HV1a'b'c, H13a2c, U5a1
MitoG5047A	5047	5046				-0.358	0.135	8.04E-03	0.025	-0.205	0.086	1.68E-02				0.017				-0.427	0.201	3.48E-02	0.027	L1b1, L3d1b, M63, M19, N2, B5a2a2a
snp6536_new683	13618	13617				0.203	0.082	1.36E-02	0.071				0.208	0.101	4.08E-02	0.065								D5b2, U5a'b
snp29998_rs41520546	12415	12414				-0.339	0.139	1.49E-02	0.024															L4a, L3e2a3, M1a3b, M54, C1c6, M27a2a, D1d, N2, P8
MitoC16272T	16272	16270				0.177	0.080	2.72E-02	0.076															L1b, L3h1a2b, M2a1, M13b2, M61a, M39c, M52a1b1, D4a8, I1c1, N11b, A2aj, H1ba, T2m, P1d1, U5, U5b2a1, K2b1a1
snp298B_new2181	5461	5460				-0.207	0.099	3.65E-02	0.049															Homoplastic[1]
MitoT3198C	3198	3197				0.156	0.079	5.00E-02	0.077															L3f1a1, L3e3b, H14b, U5a'b, U2e1a1a
snp1874C_new195	10399	10398								-0.074	0.028	7.38E-03				0.205								Homoplastic[2]
snp9711_new1384	16280	16278								-0.191	0.073	9.33E-03				0.024								Homoplastic[3]
snp2285B_new1961	4217	4216								-0.069	0.028	1.34E-02				0.199								L1b1a3b, L4b1a, M2a1b, M2c, M14, D5c, A15c1, X2b7, H1bm, H10a, R2'JT, B4a1a1a21, P4a, K1a2c
MitoA11252G	11252	11251								-0.068	0.028	1.58E-02				0.196								JT
snp1721B_new1012	15453	15452								-0.066	0.028	1.84E-02				0.195								JT, F1a3a3
snp12292_new1641	228	228								-0.099	0.046	3.32E-02				0.061								L0d1b1c, L1b1a7a, M6a2, M27a1, N5a, H7i, H13a1a6, R6a, B2a4a, R31b, U5b1g, L0d1, M45a, M23, D4f1, A16, H, J1c3c, F1a1c3, B4a1a1a9, U2b
snp4187A_new1706	2708	2706								0.046	0.022	3.98E-02				0.445								
snp4877B_new2190	5496	5495											0.451	0.209	3.21E-02	0.014								W1a, H28a2, U5a1a1
snp9329B_new2176	5427	5426											-0.536	0.260	4.03E-02	0.009								M2a3, M35a1, T2b24a, T2f1a, U2e
snp16394_rs41419549	5005	5004															-0.153	0.077	4.59E-02				0.021	L5a1a, D1j1a, D2b1a, H4
snp33532_rs28358270	9124	9123															-0.148	0.075	4.79E-02				0.022	B4a, A22, H4, H5a3a1
snp10259B_new1410	16300	16298																		-0.422	0.189	2.68E-02	0.031	L3h1a1, M7b1a1a1, M8, C1d1c1, W1d, N10b, HV0, H6a1a8, T2f1a, F3, B2g1, B4b1b, R30b1

Table 5.3: Summary of top mitochondrial SNP hits (p< 0.05) for linear GWAS tests performed on ALSPAC samples for the association of mitochondrial variants effect on mitochondrial DNA copy number variability. Data was stratified by samples extracted from whole blood or white blood cell pellets then tests were performed with or without cell count (absolute lymphocyte count and absolute neutrophil count) adjustments. rCRS BP= revised Cambridge Reference Sequence base position, SE= standard error, MAF= minor allele frequency. 1Associated with 41 clades. 2Associated with 24 clades. 3Associated with 53 clades.

5.4.4 Meta-analysis

To make UKBS and ALSPAC cohorts consistent and comparable, meta-analysis was performed on white blood cell extracted samples only. To investigate the effects of cell count adjustment on meta-analysed results, meta-analysis was performed using the following combinations:

- ALSPAC mothers extracted from WBC and all UKBS – not cell count adjusted
- ALSPAC mothers extracted from WBC and all UKBS – cell count adjusted
- ALSPAC mothers extracted from WBC and UKBS females – not cell count adjusted
- ALSPAC mothers extracted from WBC and UKBS females – cell count adjusted

Meta-analysis results are summarised in Table 5.4.

As Table 5.4 shows, mtSNP *G5046A* and *G5460A* (which both call for haplogroup W) are consistently significantly associated with mtDNA CN changes across meta-analysis, except in female only meta-analysis without cell count adjustment, whereby *G8269A* (which calls for haplogroup J1b) is the only significant mtSNP. *A3720G* (which calls for haplogroup U2) is significant in all UKBS and ALSPAC mother analysis without cell count adjustment.

Despite their significance, no meta-analysed top mtSNP results have stronger p values when combining p value power between cohorts, except *G5460A* in female only, cell count adjusted analysis. This may be due to differences in beta values between cohorts.

Irrespective, both *G5046A* and *G5460A* are consistently significant before and after meta-analysis, and have negative beta values associated with mtDNA CN. These results support the initial results reported in Figure 5.5, and show that haplogroup W individuals have a significantly lower mtDNA CN than other European mitochondrial haplogroups.

All UKBS and ALSPAC (WC) Mothers – Not Cell Count Adjusted															
CHR	*RSID*	*rCRS pos*	*allele_A*	*allele_B*	*P_value*	*BETA*	*SE*	*Q*	*P_heterogeneity*	*I2*	*ALSPAC_P*	*ALSPAC_BETA*	*ALSPAC_SE*	*UKBS_P*	*UKBS_BETA*	*UKBS_SE*
26	26_G5047A	5046	G	A	1.50E-03	-0.203	0.064	39.11	4.00E-10	97.44	9.82E-01	0.002	0.072	2.95E-12	-0.984	0.140
26	26_G5461A	5460	G	A	5.59E-03	-0.123	0.044	11.27	7.87E-04	91.13	3.71E-01	-0.045	0.050	2.12E-05	-0.408	0.096
26	26_G8270A	8269	G	A	1.56E-02	-0.130	0.054	2.75	9.72E-02	63.65	3.57E-03	-0.179	0.061	7.65E-01	0.034	0.112
26	26_A3721G	3720	A	G	5.03E-02	0.165	0.084	0.42	5.16E-01	0.00	1.29E-01	0.140	0.092	1.63E-01	0.286	0.205
All UKBS and ALSPAC (WC) Mothers – Cell Count Adjusted																
CHR	*RSID*	*rCRS pos*	*allele_A*	*allele_B*	*P_value*	*BETA*	*SE*	*Q*	*P_heterogeneity*	*I2*	*ALSPAC_P*	*ALSPAC_BETA*	*ALSPAC_SE*	*UKBS_P*	*UKBS_BETA*	*UKBS_SE*
26	26_G5047A	5046	G	A	5.12E-12	-0.805	0.117	5.32	2.11E-02	81.19	3.48E-02	-0.427	0.201	4.44E-12	-0.996	0.143
26	26_G5461A	5460	G	A	3.75E-06	-0.373	0.081	3.04	8.13E-02	67.10	1.86E-01	-0.181	0.137	2.05E-06	-0.476	0.100
UKBS Females and ALSPAC (WC) Mothers – Not Cell Count Adjusted																
CHR	*RSID*	*rCRS pos*	*allele_A*	*allele_B*	*P_value*	*BETA*	*SE*	*Q*	*P_heterogeneity*	*I2*	*ALSPAC_P*	*ALSPAC_BETA*	*ALSPAC_SE*	*UKBS_P*	*UKBS_BETA*	*UKBS_SE*
26	26_G8270A	8269	G	A	1.94E-02	-0.133	0.057	4.14	4.20E-02	75.83	3.57E-03	-0.179	0.061	2.96E-01	0.164	0.157
UKBS Females and ALSPAC (WC) Mothers – Cell Count Adjusted																
CHR	*RSID*	*rCRS pos*	*allele_A*	*allele_B*	*P_value*	*BETA*	*SE*	*Q*	*P_heterogeneity*	*I2*	*ALSPAC_P*	*ALSPAC_BETA*	*ALSPAC_SE*	*UKBS_P*	*UKBS_BETA*	*UKBS_SE*
26	26_G5047A	5046	G	A	2.58E-06	-0.667	0.142	2.83	9.23E-02	64.72	3.48E-02	-0.427	0.201	6.89E-06	-0.905	0.200
26	26_G5461A	5460	G	A	6.84E-03	-0.267	0.099	0.83	3.64E-01	0.00	1.86E-01	-0.181	0.137	1.17E-02	-0.360	0.143

Table 5.4: Top mitochondrial SNP hits (p<0.05) from meta-analysis performed between UKBS and ALSPAC cohorts, with and without cell count adjustments. CHR: chromosome. RSID: mitochondrial SNP ID. rCRS pos: revised Cambridge Reference Sequence position. Allele A: major allele. Allele B: minor allele. P_value: combined P value. SE: combined standard error. Q: Cochran's Q statistics. P_heterogeneity: p-value for heterogeneity. I2: percentage of total variation across studies that is due to heterogeneity.

5.4.5 Identifying Mitochondrial DNA Copy Number Differences between Mitochondrial Haplogroups

mtSNP genotyping GWAS analysis identified the importance of controlling for DNA source, and adjusting for cell count composition in genomic analysis to ensure consistency between cohorts. Comparison of the qPCR measurements between UKBS and ALSPAC showed minor differences in actual mtDNA CN values (section 4.4.3.1 - Figure 4.16). Therefore, to ensure the significant associations to *G5046A* and *G5460A* (defining haplogroup W) in the meta-analysis results were not an artefact of qPCR assay differences between UKBS and ALSPAC cohorts, mtDNA CN was measured in an independent, in-house control cohort (‘cohort two’) at Newcastle University, critically using the same qPCR method that was used to determine mtDNA CN in UKBS samples. Cohort two would act as a third replication cohort.

5.4.5.1 Quality Control Measures to Compare UKBS and Cohort Two Cohorts

mtDNA CN measurements were conducted, and genotyping performed on cohort two samples. Of the 1627 cohort two samples that had mitochondrial genotyping data, 1605 samples (98.64%) also had mtDNA CN determined (Figure 5.7).

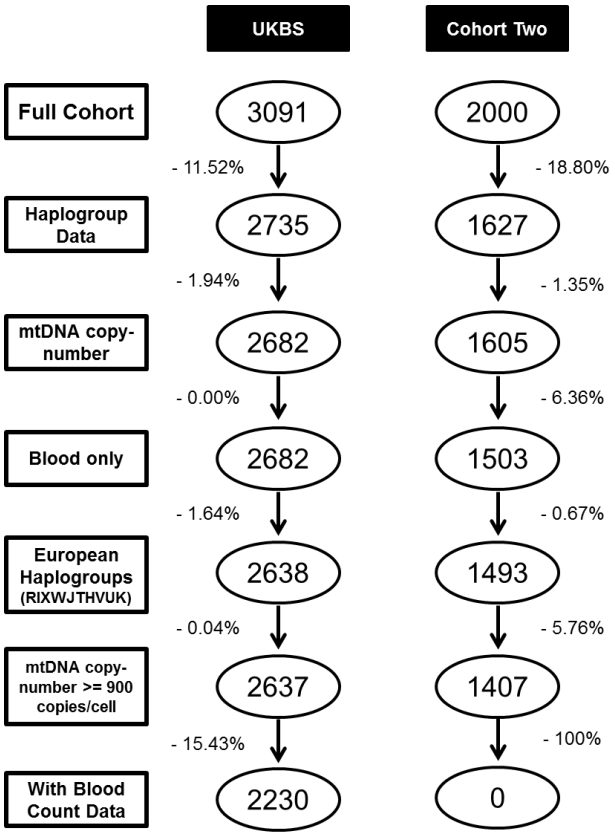


Figure 5.7: Flow diagram summarising the number of individuals in the UKBS and cohort two cohorts at each respective QC step.

Cohort two DNA was provided from blood samples and buccal samples. Therefore, to fairly compare cohorts, cohort two was filtered to contain mtDNA CN data generated from blood-extracted DNA only (n=1503). To keep my current analysis consistent with my previous analysis, only European haplogroup samples were selected (RIXWJTHVUK), removing a further 10 samples (0.67%). Of the 1493 samples remaining, samples with a mtDNA CN value greater than 900 copies/ cell were then removed (n=86, 5.76%) because they were considered outliers. This left 1407 individuals for analysis. Figure 5.7 summarises the QC steps, which were also applied to the UKBS cohort.

As UKBS DNA samples were all extracted from white blood cell pellets, DNA source did not need to be accounted for. To make data between UKBS and cohort two comparable, UKBS samples with mtDNA CN data less than or equal to 900 copies/

Mitochondrial Genetic Factors Modulating Mitochondrial DNA Copy Number

cell, and a common European haplogroup (RIXWJTHVUK) were selected (n=2637, 85.31%). Table 5.5 summarises the frequencies of each European haplogroup in the UKBS and cohort two samples, following QC steps three (Table 5.5A), and six (Table 5.5B) in Figure 5.7.

A

	UKBS		Cohort Two		
Mitochondrial Haplogroup	**N**	**%**	**N**	**%**	**Chi-sq**
R	55	2.1	6	0.4	0.0001
HV			38	2.4	
H	1149	42.8	755	47.0	0.0081
V	88	3.3	23	1.4	0.0003
JT			1	0.1	
J	316	11.8	159	9.9	0.0653
T	248	9.2	161	10	0.4282
K	226	8.4	142	8.8	0.6747
U	394	14.7	212	13.2	0.1928
W	52	1.9	30	1.9	0.9633
X	38	1.4	17	1.1	0.3860
M	15	0.6	1	0.1	0.0201
I	72	2.7	50	3.1	0.4679
L	15	0.6	8	0.5	0.9618
N1'5	14	0.5	2	0.1	0.0709
Total	**2682**	**100**	**1605**	**100**	

B

	UKBS		Cohort Two		
Mitochondrial Haplogroup	**N**	**%**	**N**	**%**	**Chi-sq**
R	55	2.1	6	0.4	0.0001
HV			35	2.5	
H	1149	43.6	678	48.2	0.0055
V	88	3.3	22	1.6	0.0014
J	316	12.0	140	10.0	0.0581
T	248	9.4	143	10.2	0.4704
K	226	8.6	117	8.3	0.8276
U	394	14.9	186	13.2	0.1497
W	52	2.0	24	1.7	0.6368
X	37	1.4	13	0.9	0.2444
I	72	2.7	43	3.1	0.6211
Total	**2637**	**100**	**1407**	**100**	

Table 5.5: Mitochondrial haplogroup frequencies in UKBS and cohort two samples. [A] Data only includes individuals who have mitochondrial DNA copy number and haplogroup data. [B] Data restricted to individuals whose DNA is from blood, is a European haplogroup and their mitochondrial DNA copy number is less than or equal to 900 copies/cell. Chi-squared tests were performed with Yates' correction.

The haplogroup frequencies observed in Table 5.5 are similar to haplogroup frequencies reported in the literature (Torroni *et al.*, 1998; Hay, 2004; Saxena *et al.*, 2006). This confirmed that the UKBS and cohort two were both representative of European populations.

5.4.5.2 Statistical Analysis of Mitochondrial Haplogroup Effect on Mitochondrial DNA Copy Number

Using IBM SPSS Statistics version 22 (IBM Corporation, 2013), haplogroups were grouped in to haplogroup versus non-haplogroup, and mtDNA CN differences between both groups were compared using ITT, and MWU tests (Figure 5.8 and Figure 5.9).

As Figure 5.8 shows, mtDNA CN is significantly lower in mitochondrial haplogroup W for both UKBS (ITT: p= 1.19E-12, MWU: p < 1.00E-04), and cohort two cohorts (ITT: p= 0.059, MWU: p= 0.008). No other mitochondrial haplogroup has significant differences in mtDNA CN compared to their respective non-haplogroups (Figure 5.9).

Figure 5.8: Mitochondrial DNA copy number (copies/cell) (y-axis) differences between UKBS and cohort two (x-axis), stratified by European mitochondrial haplogroups R, I, X, and W. Black boxes: Haplogroup. Grey boxes: Non-haplogroup. Error lines indicate standard errors. T= independent T-test. MW= Mann-Whitney U test.

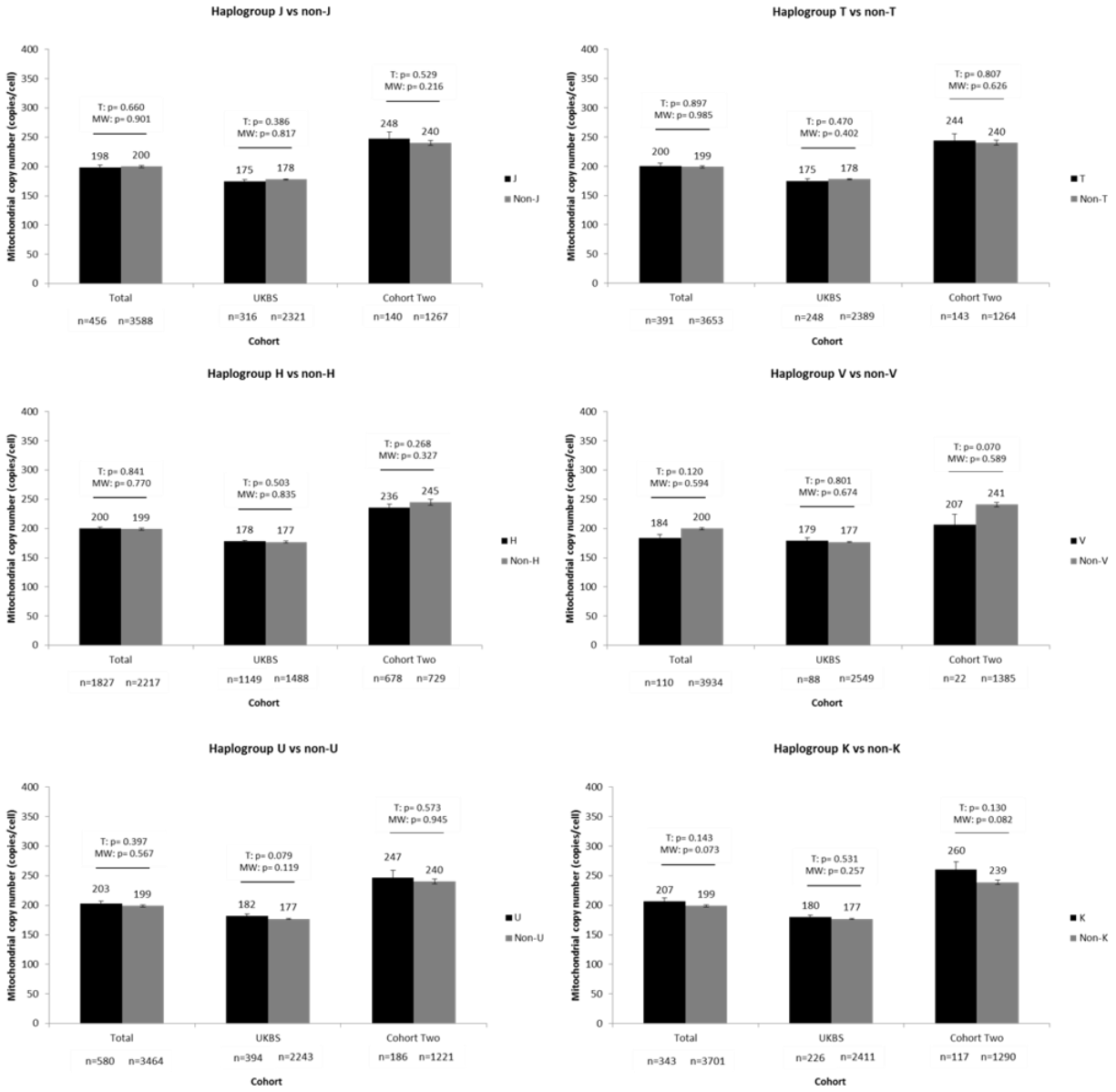


Figure 5.9: Mitochondrial DNA copy number (copies/cell) (y-axis) differences between UKBS and cohort two (x-axis), stratified by European mitochondrial haplogroups J, T, H, V, U, and K. Black boxes: Haplogroup. Grey boxes: Non-haplogroup. Error lines indicate standard errors. T= independent T-test. MW= Mann-Whitney U test.

5.4.6 Validating the Haplogroup W Effect Using an Alternative qPCR Method

In addition to biological factors contributing to mtDNA CN variability, technical errors affect mtDNA CN values (reviewed in section 3.1.2). To investigate whether the haplogroup W effect reported was due to technical limitations, I determined if any haplogroup-defining variants were located in any of the qPCR primer or probe spanning regions, as these variants may decrease their binding specificity and reduce PCR amplification or fluorescence (Gervais *et al.*, 2010).

In the UKBS cohort, genotyping variant *A3481G* at position 3480, which defines the European mitochondrial haplogroup U8b'c, was located at the final base position of the forward *MT-ND1* primer. In cohort two, one haplogroup-defining variant at position 3505 in the mtDNA was located in the *MT-ND1* qPCR amplicon region (bases 3458 to 3568), and was positioned in the centre of where the HEX probe bound (bases 3492 to 3516). This *A3505G* variant defined individuals as haplogroup W, suggesting our haplogroup W results reported may be a result of a technical artefact (Figure 5.10).

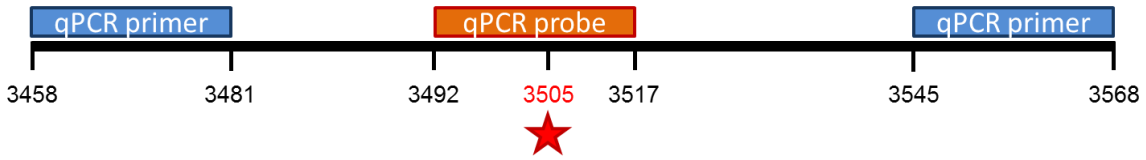


Figure 5.10: A schematic diagram summarising that qPCR DNA primers and probes potentially span across mtDNA regions which code for mitochondrial haplogroup variants. This example shows mitochondrial variant A3505G, which calls for haplogroup W, is present in the binding region for the MT-ND1-HEX qPCR probe.

Additionally, no haplogroup defining mtSNPs were located in the *MT-ND4* amplicon region in UKBS samples, and only one variant at position 11204 was located in the *MT-ND4* amplicon region (bases 11144 to 11250) in cohort two, and this was used to define individuals as mitochondrial haplogroup H. The variant was not located in the Cy5 probe region, and no mtDNA CN differences were reported in haplogroup H individuals, therefore I did not pursue altering the *MT-ND4* primers and probes.

To address the potentially technical limiting issue of the haplogroup W-defining variant being located in the centre of the *MT-ND1* fluorophore domain (Figure 5.10), an alternative mtDNA amplicon location was researched and designed (method details are in section 5.3.6).

5.4.6.1 Minor Allele Frequency of Variants around the Mitochondrial Genome

Polymorphic variants are located across the entire mtDNA molecule (Figure 5.11A), however the frequency of these variants varies between populations. To investigate regions of mtDNA which contained high frequency variants in European populations, I used a dataset where Hudson G. and colleagues downloaded 18,114 human mitochondrial genome sequences from the National Centre for Biotechnology Information (NCBI) Nucleotide database, and cleaned the data to only contain human, non-pathogenic, European and non-truncated sequences. The 9935 sequences remaining were then aligned to MUSCLE and had their mitochondrial haplogroups assigned using Haplogrep (Kloss-Brandstätter *et al.*, 2011; Weissensteiner *et al.*, 2016), leaving 2873 variants. These variants were then merged with a case-control population of 7729 individuals to perform imputation association tests (Hudson *et al.*, 2014).

Using this dataset, I looked at the MAF of each mitochondrial variant at each base position across mtDNA in a large population of people, and if any high frequency mtDNA variants (MAF >0.01) were located in any of my PCR and qPCR amplicon regions, or primer and probe-binding sites. The MAFs were then plotted and compared to see where regions of mtDNA were susceptible to variation in European populations (Figure 5.11A). Figure 5.11B summarises the frequency of MAF variation across mtDNA, where the majority of mtDNA variants have a MAF of about 0.001.

As Figure 5.11C and Table 5.6 show, two highly frequent (MAF >0.01) variants are located in both the primer and probe binding regions of *MT-ND1*, but no highly frequent variants are located in the *MT-ND4* amplicon.

An amplicon region spanning from bases 2911 to 3036, in *MT-RNR2*, which codes for the mitochondrial 16S ribosomal subunit, was identified as a suitable region to quantify instead of *MT-ND1. MT-RNR2* neighboured *MT-ND1*, and was therefore adjacent to *MT-ND4* if any mtDNA deletions needed to be determined, but also this region contained no high frequency variants that spanned primer or probe binding sites. Additionally, no haplogroup-defining variants were located in this *MT-RNR2* amplicon region for cohort two, and only *G3012A* at mtDNA position 3010 was present in UKBS genotyping. *G3012A* called for haplogroups H, J and U, which were all reported to have no mtDNA CN changes. Therefore I chose to use this as an alternative qPCR site to *MT-ND1* to re-quantify mtDNA CN.

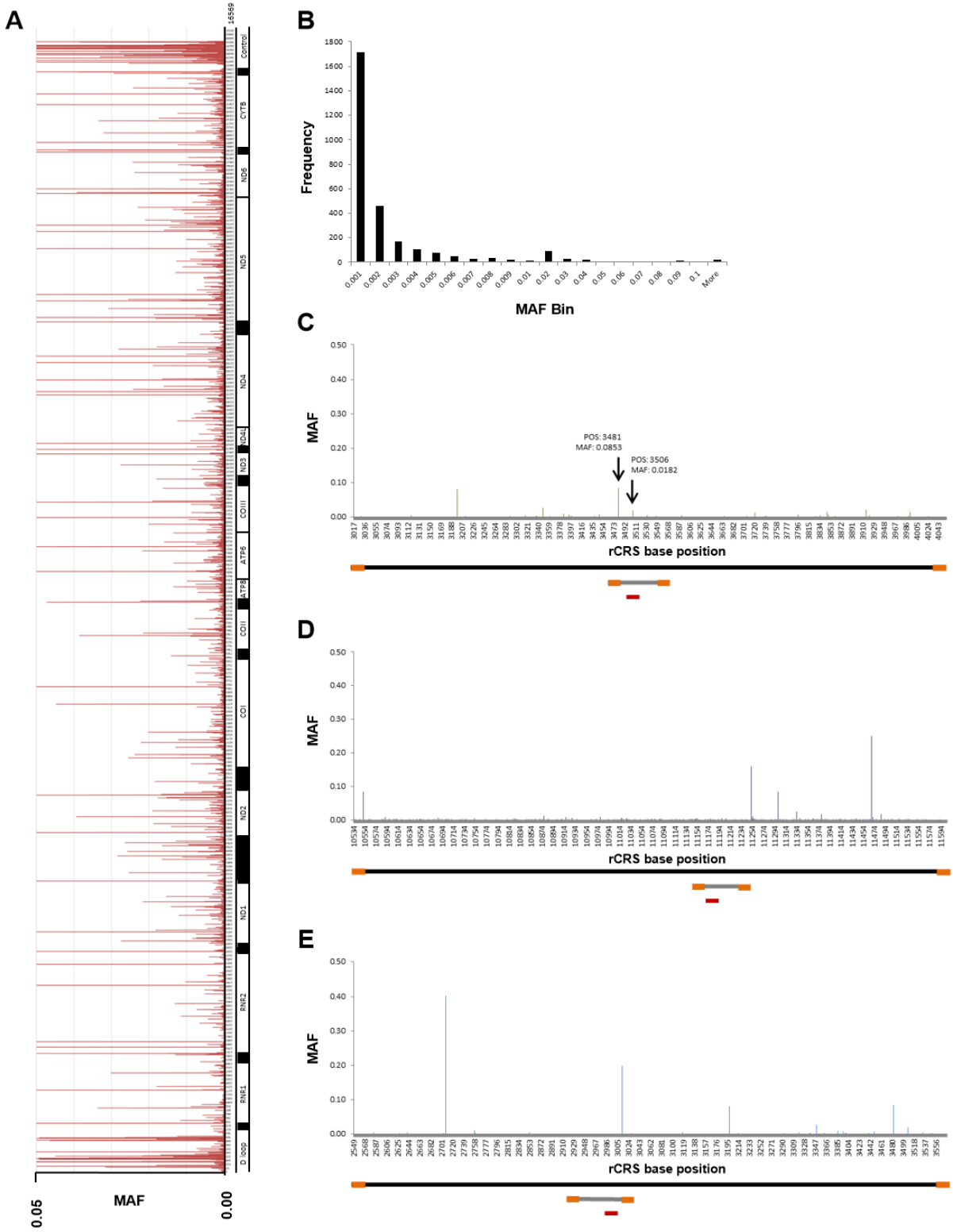


Figure 5.11: Graphs summarising the MAF of mitochondrial variants across [A] the mitochondrial genome, [C] MT-ND1, [D] MT-ND4 and [E] MT-RNR2 (16S). In [C] – [E] the black lines represent standard template DNA amplicon regions, grey lines indicate qPCR amplicon regions and red lines indicate probe binding regions. Orange boxes indicate primer binding sites. [B] Frequency distribution summarising the number of variants with a specific MAF across the whole mitochondrial genome in 7729 CEU people (adapted from (Hudson et al., 2014).

MT-ND1 qPCR amplicon positions	MAF	*MT-ND4* qPCR amplicon positions	MAF	*MT-RNR2* qPCR amplicon positions	MAF	Bristol D-loop qPCR amplicon positions	MAF
3458	0.0000	11144	0.0013	2911	0.0000	317	0.0000
3459	0.0000	11145	0.0006	2912	0.0000	318	0.0005
3460	0.0004	11146	0.0000	2913	0.0000	319	0.0000
3461	0.0009	11147	0.0000	2914	0.0000	320	0.0000
3462	0.0000	11148	0.0004	2915	0.0000	321	0.0019
3463	0.0000	11149	0.0000	2916	0.0000	322	0.0065
3464	0.0000	11150	0.0005	2917	0.0000	323	0.0003
3465	0.0000	11151	0.0014	2918	0.0000	324	0.0005
3466	0.0000	11152	0.0016	2919	0.0000	325	0.0000
3467	0.0000	11153	0.0044	2920	0.0000	326	0.0000
3468	0.0000	11154	0.0000	2921	0.0000	327	0.0003
3469	0.0000	11155	0.0000	2922	0.0000	328	0.0000
3470	0.0000	11156	0.0009	2923	0.0000	329	0.0012
3471	0.0000	11157	0.0000	2924	0.0000	330	0.0000
3472	0.0000	11158	0.0000	2925	0.0000	331	0.0009
3473	0.0000	11159	0.0000	2926	0.0000	332	0.0004
3474	0.0000	11160	0.0000	2927	0.0000	333	0.0000
3475	0.0000	11161	0.0000	2928	0.0000	334	0.0000
3476	0.0000	11162	0.0000	2929	0.0000	335	0.0000
3477	0.0000	11163	0.0000	2930	0.0000	336	0.0003
3478	0.0000	11164	0.0000	2931	0.0000	337	0.0004
3479	0.0000	11165	0.0000	2932	0.0000	338	0.0000
3480	0.0000	11166	0.0000	2933	0.0000	339	0.0005
3481	0.0853	11167	0.0000	2934	0.0000	340	0.0004
3482	0.0000	11168	0.0012	2935	0.0000	341	0.0000
3483	0.0000	11169	0.0000	2936	0.0000	342	0.0012
3484	0.0016	11170	0.0000	2937	0.0000	343	0.0000
3485	0.0000	11171	0.0000	2938	0.0000	344	0.0000
3486	0.0000	11172	0.0000	2939	0.0000	345	0.0000
3487	0.0000	11173	0.0016	2940	0.0000	346	0.0000
3488	0.0000	11174	0.0000	2941	0.0000	347	0.0072
3489	0.0000	11175	0.0000	2942	0.0000	348	0.0000
3490	0.0000	11176	0.0000	2943	0.0000	349	0.0000
3491	0.0000	11177	0.0032	2944	0.0000	350	0.0000
3492	0.0000	11178	0.0019	2945	0.0000	351	0.0000
3493	0.0000	11179	0.0000	2946	0.0000	352	0.0000
3494	0.0000	11180	0.0000	2947	0.0000	353	0.0000
3495	0.0000	11181	0.0000	2948	0.0000	354	0.0000
3496	0.0000	11182	0.0000	2949	0.0000	355	0.0000
3497	0.0000	11183	0.0004	2950	0.0000	356	0.0000
3498	0.0000	11184	0.0000	2951	0.0000	357	0.0000
3499	0.0004	11185	0.0000	2952	0.0000	358	0.0000
3500	0.0000	11186	0.0000	2953	0.0000	359	0.0000
3501	0.0000	11187	0.0000	2954	0.0000	360	0.0000
3502	0.0000	11188	0.0000	2955	0.0000	361	0.0000
3503	0.0000	11189	0.0000	2956	0.0000	362	0.0000
3504	0.0000	11190	0.0000	2957	0.0003	363	0.0000
3505	0.0008	11191	0.0000	2958	0.0000	364	0.0000
3506	0.0182	11192	0.0016	2959	0.0000	365	0.0000
3507	0.0000	11193	0.0000	2960	0.0000	366	0.0000
3508	0.0000	11194	0.0000	2961	0.0000	367	0.0000
3509	0.0000	11195	0.0000	2962	0.0000	368	0.0000
3510	0.0004	11196	0.0000	2963	0.0000	369	0.0000
3511	0.0000	11197	0.0000	2964	0.0000	370	0.0004
3512	0.0000	11198	0.0043	2965	0.0000	371	0.0000
3513	0.0000	11199	0.0000	2966	0.0000	372	0.0000
3514	0.0003	11200	0.0000	2967	0.0000	373	0.0000
3515	0.0000	11201	0.0000	2968	0.0000	374	0.0003
3516	0.0000	11202	0.0000	2969	0.0000	375	0.0006
3517	0.0000	11203	0.0000	2970	0.0000	376	0.0034
3518	0.0000	11204	0.0000	2971	0.0000	377	0.0000
3519	0.0000	11205	0.0048	2972	0.0000	378	0.0000
3520	0.0000	11206	0.0000	2973	0.0000	379	0.0000
3521	0.0000	11207	0.0000	2974	0.0000	380	0.0000
3522	0.0000	11208	0.0000	2975	0.0000	381	0.0000
3523	0.0000	11209	0.0000	2976	0.0000		
3524	0.0000	11210	0.0000	2977	0.0000		
3525	0.0000	11211	0.0000	2978	0.0000		
3526	0.0012	11212	0.0000	2979	0.0000		
3527	0.0000	11213	0.0000	2980	0.0000		
3528	0.0000	11214	0.0000	2981	0.0000		
3529	0.0000	11215	0.0000	2982	0.0000		
3530	0.0000	11216	0.0000	2983	0.0000		
3531	0.0000	11217	0.0000	2984	0.0000		
3532	0.0044	11218	0.0000	2985	0.0000		
3533	0.0000	11219	0.0000	2986	0.0000		
3534	0.0003	11220	0.0000	2987	0.0000		
3535	0.0000	11221	0.0000	2988	0.0000		
3536	0.0000	11222	0.0003	2989	0.0000		
3537	0.0000	11223	0.0000	2990	0.0000		
3538	0.0005	11224	0.0000	2991	0.0000		
3539	0.0000	11225	0.0000	2992	0.0000		
3540	0.0000	11226	0.0000	2993	0.0000		
3541	0.0005	11227	0.0000	2994	0.0000		
3542	0.0000	11228	0.0000	2995	0.0000		
3543	0.0000	11229	0.0000	2996	0.0000		
3544	0.0003	11230	0.0000	2997	0.0000		
3545	0.0000	11231	0.0000	2998	0.0000		
3546	0.0000	11232	0.0000	2999	0.0000		
3547	0.0009	11233	0.0000	3000	0.0000		
3548	0.0022	11234	0.0013	3001	0.0000		
3549	0.0006	11235	0.0000	3002	0.0000		
3550	0.0010	11236	0.0000	3003	0.0000		
3551	0.0000	11237	0.0000	3004	0.0000		
3552	0.0000	11238	0.0000	3005	0.0000		
3553	0.0000	11239	0.0000	3006	0.0000		
3554	0.0000	11240	0.0000	3007	0.0000		
3555	0.0000	11241	0.0000	3008	0.0000		
3556	0.0000	11242	0.0000	3009	0.0000		
3557	0.0000	11243	0.0014	3010	0.0000		
3558	0.0000	11244	0.0000	3011	0.0000		
3559	0.0000	11245	0.0000	3012	0.1991		
3560	0.0000	11246	0.0000	3013	0.0000		
3561	0.0000	11247	0.0000	3014	0.0000		
3562	0.0000	11248	0.0000	3015	0.0000		
3563	0.0000	11249	0.0000	3016	0.0000		
3564	0.0000	11250	0.0000	3017	0.0000		
3565	0.0004			3018	0.0000		
3566	0.0005			3019	0.0000		
3567	0.0000			3020	0.0000		
3568	0.0003			3021	0.0000		
				3022	0.0000		
				3023	0.0000		
				3024	0.0000		
				3025	0.0000		
				3026	0.0000		
				3027	0.0000		
				3028	0.0000		
				3029	0.0026		
				3030	0.0000		
				3031	0.0000		
				3032	0.0000		
				3033	0.0000		
				3034	0.0000		
				3035	0.0000		
				3036	0.0000		

Table 5.6: Breakdown of the MAF of mitochondrial variants located across qPCR amplicon sites in mtDNA (MT-ND1, MT-ND4, MT-RNR2 and D-loop). Orange indicates primer binding regions. Pink indicates probe binding regions. Blue indicates variants with MAF>0.00 and red indicates variants with MAF>0.01.

5.4.6.2 Replicating the Mitochondrial Haplogroup W Effect using an Alternative qPCR Method

5.4.6.2.1 Cohort One - UKBS

To validate if the significantly lower mtDNA CN reported in haplogroup W individuals was real, and not due to technical artefacts, 108 UKBS samples (n=53 haplogroup W and n=55 non-haplogroup W) were re-measured using the alternative qPCR method which quantified *MT-RNR2* instead of *MT-ND1* (details in section 5.3.6.1). 105 samples (97.22%) had mtDNA CN values calculated. These samples were then matched with their previous mtDNA CN values, which were calculated using the relative proportion of *MT-ND1* to *B2M*, and were grouped in to haplogroup W versus non-haplogroup W groups. ITT and MWU tests were performed on each group to compare mtDNA CN differences between haplogroups, and the potential effect of the qPCR method applied on mtDNA CN values (Figure 5.12).

Using the original qPCR assay measuring *B2M*, *MT-ND1,* and *MT-ND4*, initial results reported mtDNA CN to be significantly lower (p= 9.64E-10) in haplogroup W individuals (μ= 130.72 copies/cell, ±37.15 s.d) than in non-haplogroup W individuals (μ= 192.56 copies/cell, ±53.09 s.d) (Figure 5.12A). These results were replicated when using the alternative qPCR method, measuring *B2M*, *MT-RNR2,* and *MT-ND4* because mtDNA CN was still significantly lower (p= 0.026) in haplogroup W individuals (μ= 200.31 copies/cell, ±67.77 s.d) than in non-haplogroup W individuals (μ= 241.77 copies/cell, ±113.88 s.d) (Figure 5.12B).

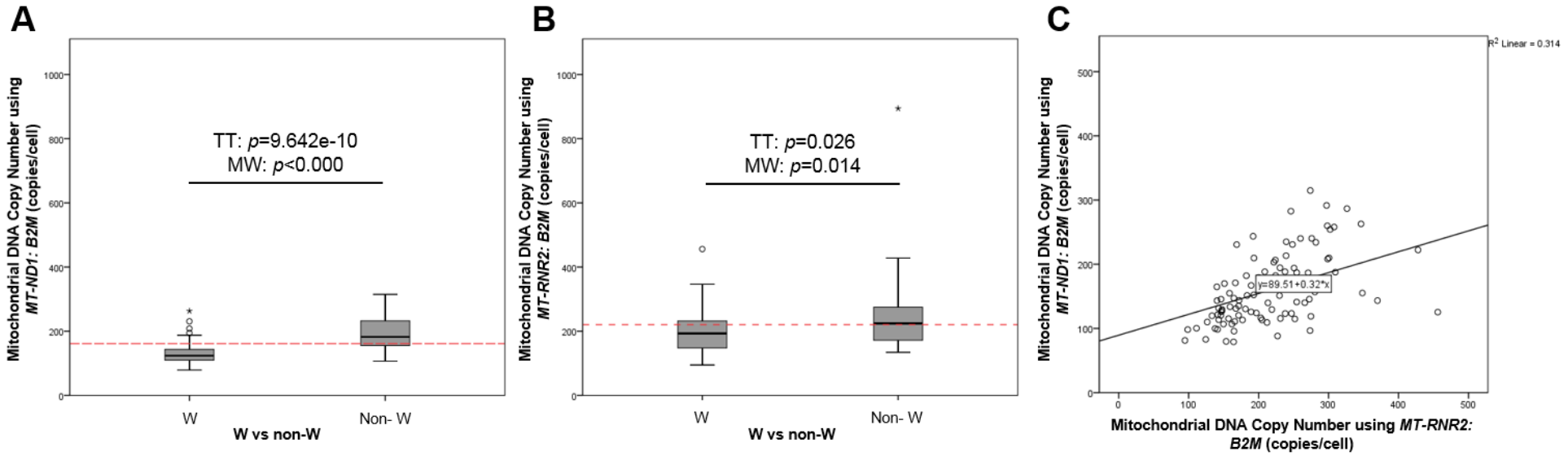


Figure 5.12: Graphs demonstrating the difference in mitochondrial DNA copy number between haplogroup W and non-haplogroup W individuals in UKBS samples. [A] Uses data from the original B2M, MT-ND1, and MT-ND4 qPCR assay (n=52 haplogroup W, n=51 non-haplogroup W). [B] Uses data from the alternative B2M, MT-RNR2, MT-ND4 qPCR assay (n=53 haplogroup W and n=52 non-haplogroup W). [C] Compares data from both qPCR assays. Red dotted line indicates population mean.

5.4.6.2.2 Cohort Two – In house ‘Cohort Two’

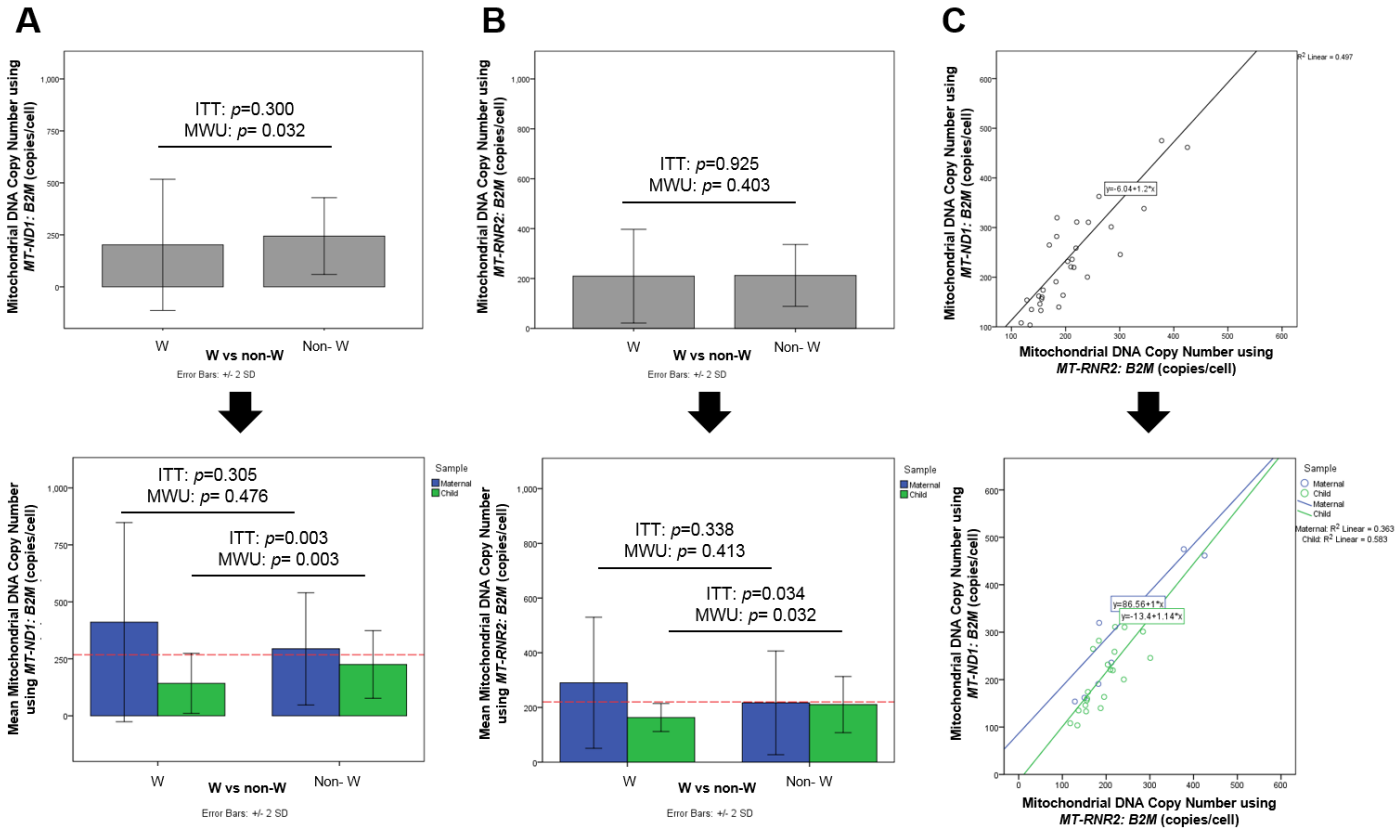
Acting as an independent, replicate study, 40 samples (n=18 haplogroup W and n=22 non-haplogroup W) from cohort two also had their mtDNA CN measured again using the alternative qPCR method (methods are detailed in section 5.3.6.1). 10 of these samples were from mothers and 30 were from children.

Following analysis, and mtDNA CN generation, each sample was grouped by haplogroup, and ITT and MWU tests were performed (Figure 5.13).

Initial ITT and MWU tests performed between haplogroup W and non-haplogroup W individuals for the alternative qPCR assay (using *MT-RNR2: B2M*) (Figure 5.13B) suggested there was no significant difference in mtDNA CN between groups (ITT: p= 0.925, MWU: p= 0.403). However, when groups were stratified by mothers and children, mtDNA CN was significantly lower in haplogroup W children (μ= 163.28 copies/cell, ±25.57 s.d versus μ= 210.51 copies/cell, ±51.38 s.d) (ITT: p= 0.034, MWU: p= 0.032), but not in mothers (μ= 290.23 copies/cell, ±119.99 s.d versus μ= 216.91 copies/cell, ±94.73 s.d) (ITT: p= 0.308, MWU: p= 0.413).

Linear regression analysis also reported that there was a positive correlation between mtDNA CN values generated using the original versus the alternative qPCR method in both UKBS analysis (β= 0.35, r^2= 0.114) (Figure 5.12C), and cohort two analysis (β= 0.705, r^2= 0.497) (Figure 5.13C).

Overall, these tests confirm that mtDNA CN is significantly lower in mitochondrial haplogroup W individuals, and it is not due to qPCR technical artefacts.



D

N	*MT-ND1: B2M* assay		*MT-RNR2: B2M* assay	
Cohort Two	**Haplogroup W**	**Non-haplogroup W**	**Haplogroup W**	**Non-haplogroup W**
All	18	22	11	20
Mothers	4	6	4	5
Children	14	16	7	15

Figure 5.13: Graphs demonstrating the difference in mitochondrial DNA copy number between haplogroup W and non-haplogroup W individuals in cohort two samples then stratified by children and mothers. [A] Uses data from the original B2M, MT-ND1, and MT-ND4 qPCR assay. [B] Uses data from the alternative B2M, MT-RNR2, and MT-ND4 qPCR assay. [C] Compares data from both qPCR assays. Red dotted line indicates population mean. Error bars indicate +/- 2 standard deviations. [D] Summarises the number of mothers and children in each mitochondrial haplogroup category.

5.5 Discussion

This is the first, and largest, study to investigate the effect of mitochondrial variants on mtDNA CN changes in healthy European cohorts. Using three, independent cohorts, I have performed two mtDNA GWAS using mtDNA CN as a continuous phenotype, identifying mtDNA variants associated with mtDNA CN changes. Meta-analysis has strengthened the power of this study by identifying associations in up to 5921 unrelated, European individuals. An in-house control cohort has further supported the results from my meta-analysis, by confirming that mtDNA CN is significantly lower in mitochondrial haplogroup W individuals, compared to any other European haplogroup. Using an alternative mtDNA reference site to calculate mtDNA CN also validated that results reported are not due to technical artefacts.

5.5.1 Haplogroup W Effect

Mitochondrial variants *G5460A* and *G5046A*, which were used to define haplogroup W, were consistently top GWAS hits in both UKBS and ALSPAC cohorts GWAS analysis, with or without blood count adjustments. Both variants were negatively correlated to mtDNA CN, as initially indicated in Figure 5.5. The individual GWAS for ALSPAC samples (Table 5.3) reported many more overall top hits than UKBS GWAS (Table 5.2); however meta-analysis showed consistent hits in both variants as power increased. The differences reported in ALSPAC samples may have been a result of population heterogeneity, as indicated by the significant values in the P_heterogeneity column in Table 5.4. However, these results are not reliable because mtSNP MAFs strongly correlated between ALSPAC and UKBS before analysis (Figure 5.6) to ensure there was no population heterogeneity. Additionally, P_heterogeneity is calculated using Cochran's Q test (Q) (StatsDirect Ltd, 2013), and Hoaglin argues that many meta-analyses report Cochran's Q test to assess heterogeneity, and these authors cite work by W. G. Cochran, without realizing that Cochran deliberately did not use Q itself to test for heterogeneity (Hoaglin, 2016). This suggests that P_heterogeneity scores may not be a viable indicator of differences between cohorts, and ALSPAC and UKBS cohorts were not heterogenic, and were comparable. Furthermore, as Table 5.5 shows, the frequency of haplogroup W in both UKBS and cohort two cohorts was similar, and agreed with the literature (Saxena *et al.*, 2006). This shows that the consistently significantly lower mtDNA CN reported in haplogroup W in these cohorts is not a result of differences in haplogroup frequencies between the cohorts.

Additionally, overall haplogroup frequencies between the UKBS and cohort two cohorts were similar, despite significant differences being reported in haplogroups R, H, V, and M between cohorts (p <0.05) when chi-squared tests with Yates' correction (designed to improve the chi-square approximation) were performed (Table 5.5A). These significant differences were likely due to UKBS samples being assigned to their mitochondrial haplogroup using 138 SNPs from SNP genotyping data, whereas cohort two used sequencing data which had more variants to define haplogroups more accurately. The overall haplogroup frequencies between cohorts were in line with the literature though (Torroni *et al.*, 1996; Torroni *et al.*, 2000), and therefore UKBS and cohort two were representative of European populations.

In Europe, mitochondrial haplogroup W is most prevalent in North-eastern countries such as Finland (9.5%), Hungary (5%), Latvia (4%), Macedonia (4%), and Belarus (3.5%) (Hay, 2004). Haplogroup WIX individuals were more common than haplogroup J in an aged Finnish population (Niemi *et al.*, 2003), discrediting the original reports suggesting haplogroup J is associated with longevity (Rose *et al.*, 2001; Dato *et al.*, 2004). However Niemi's study may be a reflection of the more prevalent haplogroup W in the Finnish population than the rest of Europe (Hay, 2004).

Furthermore, a study conducted across 1000 LHON-affected families in Finland reported that the frequency of LHON associated mtDNA mutations (*G3460A, G11778A, and T14484C*) was much lower in their population than other Northern European, UK, and Australian families, and that affected females had a higher incidence of affected offspring compared to unaffected females (Puomila *et al.*, 2007). Mitochondrial haplogroups were not reported in this study, but the higher prevalence of haplogroup W in the Finnish population could be contributing to the higher incidence of LHON being transmitted to affected mother's offspring if their mtDNA CN was significantly lower as a result of their mitochondrial haplogroup. This theory would support Giordano's study where they reported higher mtDNA CN protect against LHON penetrance (Giordano *et al.*, 2014).

Both the *G5046A* and *G5460A* mtSNPs are non-synonymous, and result in amino acid changes from a valine to isoleucine (codon 193), and from an alanine to a threonine (codon 331) respectively in *MT-ND2* (Brandon *et al.*, 2005; Lott *et al.*, 2013). Gomez-Duran and colleagues reported differences in OXPHOS capacities

between haplogroups (Gomez-Duran *et al.*, 2010), and Ruiz-Pesini and colleagues reported differences in OXPHOS capacities dependent on reaction coupling efficiencies to generate ATP (Ruiz-Pesini *et al.*, 2004). If changes of amino acid are introduced to the *MT-ND2* subunit during translation in haplogroup W individuals, this could be affecting the OXPHOS coupling strength, and potentially changing the overall OXPHOS capacity of haplogroup W mitochondria. A pharmaceutical study in HIV patients support this theory, as haplogroup W individuals reported to have the greatest increase in extremity fat when being treated with antiretroviral treatments (Hulgan *et al.*, 2011), and antiviral therapies are reported to affect mtDNA CN (Payne *et al.*, 2011), which could introduce a 'double-hit' insult to metabolic efficiency of haplogroup W individuals to cause fat accumulation.

Unlike nDNA, mtDNA is predominantly unmethylated (Groot and Kroon, 1979), and the inability to establish a mtDNA CN set point in cells can be lethal, especially during development (Dickinson *et al.*, 2013; Kelly *et al.*, 2013b). Embryonic stem cell models with the same nDNA background, but different mtDNA backgrounds have slightly different differential gene expression patterns (Kelly *et al.*, 2013a), as well as expression levels of mtDNA replication gene *POLG-A* (Kelly *et al.*, 2012). This suggests that mitochondrial haplogroups might alter the methylation profiles on nDNA, which has downstream effects on mitochondrial replication and translation rates (St. John, 2016).

In vertebrates, CpG sites are normally associated with gene promotor regions, and genes that are often expressed in tissues contain many CpG sites, normally forming CpG islands (Saxonov *et al.*, 2006). This is because CpGs are methylation sites, and the binding of methyl groups to cytosine nucleotides causes displacement of transcription factors, and attracts gene silencing and chromatin compacting proteins (Jaenisch and Bird; Kim *et al.*, 2003; Fazzari and Greally, 2004). Despite the importance of CpG sites in development (Singal and Ginder, 1999), nDNA contains lower CpG levels than expected; with total nDNA containing 25% CpG sites, 45% located in exons and 72% located to promotor regions (Saxonov *et al.*, 2006; Chinnery *et al.*, 2012). What's more is mtDNA contains lower levels of CpG sites than expected, which is believed to be a result of natural selection. However, there are a high frequency of CpG sites in the O_L site, and in between the haplogroup W defining SNPs *G5460A* and *G5046A* in *MT-ND2*, of which significantly lower mtDNA CN were associated. It is possible to hypothesise that these sites might both be affecting

Mitochondrial Genetic Factors Modulating Mitochondrial DNA Copy Number

mtDNA replication (Figure 5.14) (Chinnery *et al.*, 2012). Similar to the D-loop region, where triple-stranded structures form and interrupt the binding of mtSSB and TFAM (Takamatsu *et al.*, 2002); the formation of these methylation sites near the O_L may be interrupting mtDNA replication. Additionally, Samuels and colleagues use NGS data to report recurrent, tissue specific mtDNA mutations across several unrelated individuals. These mutations are close to replication sites of mtDNA, but not in the coding regions, and further suggest they may be affecting mtDNA replication (Samuels *et al.*, 2013).

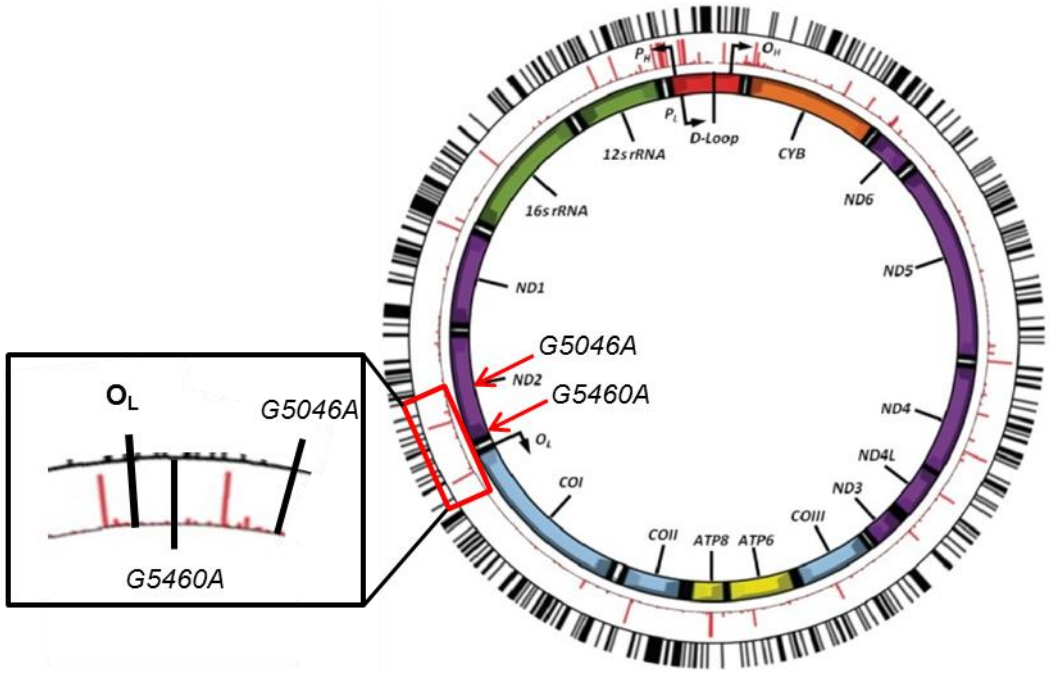


Figure 5.14: Schematic representation of mtDNA and the distribution of CpG sites relative to mitochondrial haplogroup defining polymorphisms. Haplogroup W defining variants G5460A and G5046A are labelled in MT-ND2 region. Image adapted from (Chinnery et al., 2012).

Further functional studies would need to be performed in haplogroup W mitochondria to investigate the molecular mechanisms regulating mtDNA replication, as well as understand the OXPHOS capacities for these mitochondria to operate with significantly lower mtDNA CN levels. Population studies may be difficult in haplogroup W individuals because they are a rare European haplogroup, however use of transfected cybrid cell lines with haplogroup variants, or mutants in *MT-ND2* may be useful to understand biogenesis differences (Pye *et al.*, 2006).

5.5.2 Blood Composition Differences between Haplogroups

mtDNA CN was previously reported to significantly change relative to blood composition cell counts (section 3.3.5 - Figure 3.14 and Table 3.4). To ensure the mitochondrial haplogroup W effect was not a result of differences in blood cell composition in UKBS and ALSPAC samples, I investigated whether there were any differences in blood cell composition relative to mitochondrial haplogroup using UKBS samples.

As described in section 2.3, UKBS DNA samples were extracted from blood donations. Full blood count (FBC) data was also generated using a Beckman-Coulter GenS (Gieger *et al.*, 2011), and both cell counts and proportions were measured accordingly. FBC data was available for 2184 non-haplogroup W and 46 haplogroup W individuals, and platelet count data was available for 2172 non-haplogroup W and 45 haplogroup W individuals. Using IBM SPSS Statistics version 22 (IBM Corporation, 2013), ITT and MWU tests were conducted (Figure 5.15).

As Figure 5.15 shows, there is no significant difference in any blood cell composition in haplogroup W versus non-W haplogroup individuals. This further supports that mtDNA CN is significantly lower in haplogroup W individuals due to genetic differences, and not blood composition.

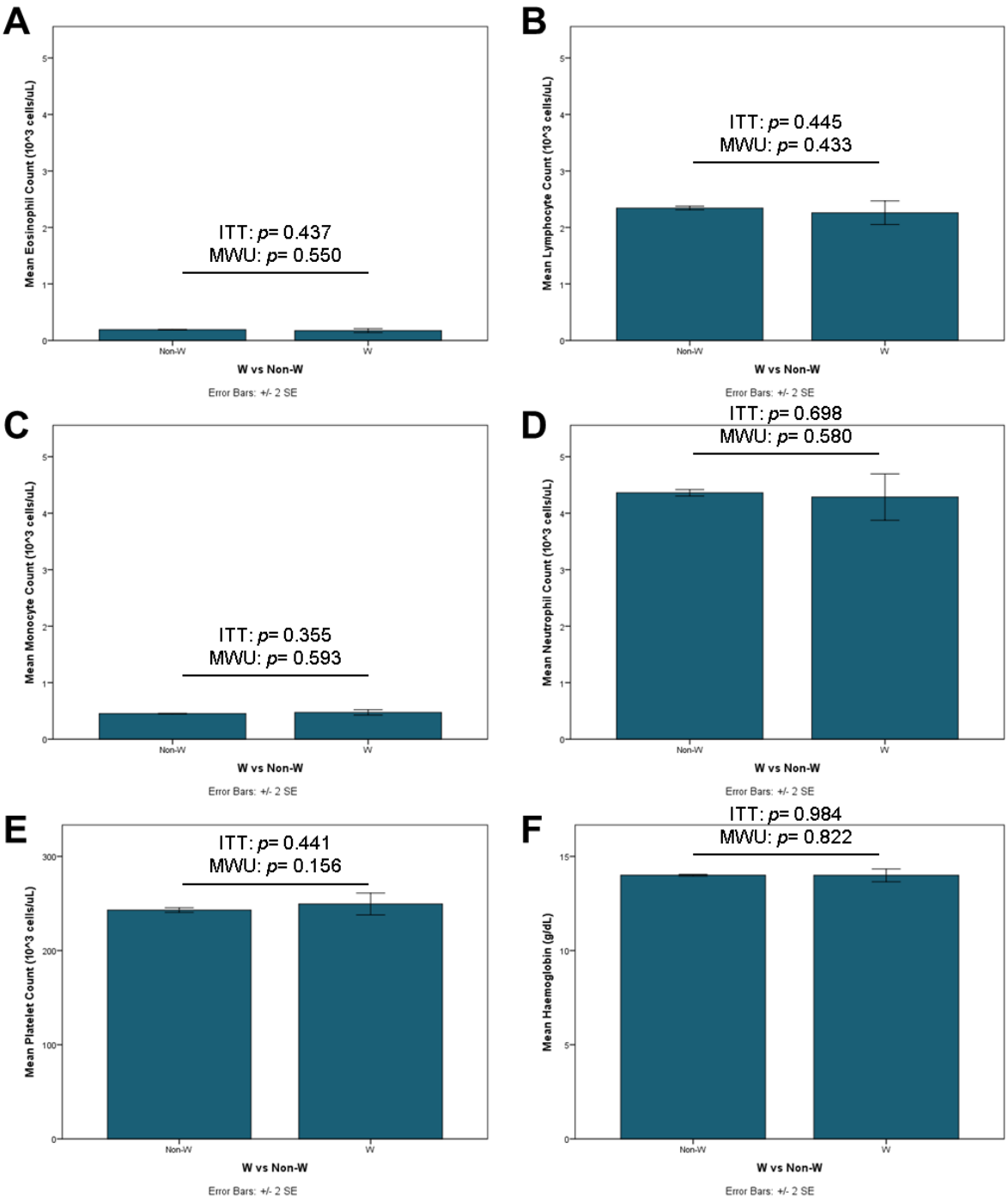


Figure 5.15: Differences in [A] eosinophil count, [B] lymphocyte count, [C] monocyte count, [D] neutrophil count, [E] platelet count and [F] haemoglobin between haplogroup W (n=52) and non-haplogroup W individuals (n=2585) in UKBS cohort.. Error bars indicate +/- 2 standard errors. ITT= independent T-test. MWU= Mann- Whitney U test.

5.5.3 Study Limitations

5.5.3.1 Replicating the Haplogroup W Effect

The significantly lower mtDNA CN reported in haplogroup W individuals compared to non-haplogroup W individuals may not replicate in cohort two mothers using the alternative qPCR method because the sample size in cohort two is very small (haplogroup W: n=4, non-haplogroup W: n=6) (Figure 5.13D), reporting larger deviations between results. Larger sample sizes reduce s.d values, making differences more obvious, as indicated in UKBS haplogroup analysis (Figure 5.8). To identify if the haplogroup W effect is real in mothers and children in cohort two analysis, as originally reported in Figure 5.8, more samples need to be measured to minimise the deviation in results reported using the alternative qPCR method.

To support this logic, mtDNA CN values measured using the original qPCR method reported a significantly lower mtDNA CN in haplogroup W individuals than non-W individuals, however this effect was lost when comparing the same, but very few samples in Figure 5.13A. Using the original qPCR method, haplogroup W was significantly lower in cohort two samples, irrespective of whether the samples were from mothers or children (Figure 5.8).

Additionally, cohort two DNA quality may be poor because very little sample was available to perform qPCR on (approximately 8 µl), and smaller DNA volumes are more sensitive to the freeze-thaw process (Krajden *et al.*, 1999). This is could explain the large s.d values generated for mothers in both Figure 5.13A and B.

Furthermore, details about the source of cohort two DNA was not available. As indicated in the ALSPAC analysis, factors such as DNA source and tissue type need to be the same to make data comparable. If cohort two mothers were extracted from blood, and cohort two children from white cell pellets, this may explain the inconsistent mtDNA CN results reported when using the alternative qPCR method. However, the poor DNA quality, and fewer samples is more likely the cause of differences.

5.5.3.2 qPCR Methods

In addition to potential genetic differences, ALSPAC GWAS results may not have replicated exactly to UKBS because of phenotype differences. Bristol used SYBR® green to measure mtDNA CN in their samples, and SYBR® green is unspecific as it

binds to any double stranded DNA. Additionally, the size of the mtDNA reference amplicon region in Bristol's qPCR method was very small (65 bp) compared to our qPCR method (Table 5.6). Amplifying smaller regions of mtDNA, using a less specific qPCR probe which could introduce inaccurate mtDNA quantification, which would result in mtDNA CN differences between our cohorts. Future work would need to ensure that similar qPCR methods were used to calculate mtDNA CN, or respective phenotypes, to make meta-analysis more accurate.

This study has identified the importance of good qPCR design to calculate mtDNA CN. Similar to Lemmon and colleagues work, where they reported false positive TaqMan™ probe and primer binding sites across 112 bacterial and viral genome signatures as a result of poor probe and primer designs (Lemmon and Gardner, 2008). I have explored the limitations of reference mtDNA regions in qPCR, and have identified that both probe and primer sites should be designed away from highly polymorphic regions, and ideally away from haplogroup defining variants when conducting mitochondrial haplogroup analysis. Measuring several mtDNA regions may also be beneficial to normalise to each other before calculating mtDNA relative to nDNA. This would ensure full, circular mtDNA was being quantified, and not linear molecules. Miller and colleagues used a plasmid in their qPCR analysis to accurately quantify mtDNA CN in their samples, which may also offer consistency between assay repeats and sample repeats (Miller *et al.*, 2003). Several studies have also suggested droplet-digital PCR (ddPCR) is more accurate for mtDNA CN determination. This is because DNA in ddPCR is diluted so that, prior to amplification, each compartment (or droplet) contains a single piece of DNA. ddPCR uses thousands to millions of reaction compartments per assay, and also uses florescence probes for detection, which enhances sensitivity and increases precision (Henrich *et al.*, 2012; Belmonte *et al.*, 2016; Ye *et al.*, 2017).

Finally, heteroplasmy is known to effect mtDNA genotype segregation, especially during the mitochondrial bottleneck during embryogenesis and development (Cree *et al.*, 2008b). This suggests it would be beneficial to determine heteroplasmy levels in samples as well as mtDNA CN, to investigate whether mtDNA CN changes are a result of heteroplasmy compensation.

This study is the first study to investigate mtDNA CN changes across three large, unrelated, healthy, European cohorts. Despite the thousands of publications

reporting mtDNA CN changes or mitochondrial haplogroup associations to disease susceptibility, progression, and severity, no study has identified ‘normal’ mtDNA CN levels in European populations between mitochondrial haplogroups. Therefore, this study acts both as a unique discovery as well as a reference for future work. For the first time, I have identified that mtDNA CN changes are specific to a rare, European haplogroup, and that these differences are not a result of technical or biological limitations. Furthermore I have identified the importance of good qPCR design, and ongoing work needs to further understand the biological mechanisms contributing to the results reported in this work.

Chapter 6 Mitochondrial DNA Copy Number in Pre- and Post-Menopausal Women

6.1 Introduction

6.1.1 Menopause

Menopause is when women stop menstruating, and is a natural part of ageing. In the UK, 1% of women experience menopause before 40 years old (Coulam *et al.*, 1986; NHS, 2017), however the average age of onset is typically between 45 and 55 years (Cramer *et al.*, 1995; Torgerson *et al.*, 1997), with over 75% of women undergoing menopause before the age of 50 (Broekmans *et al.*, 2004) (Figure 6.1). Onset of menopause is multifactorial, and genetic factors have been suggested to play a major role, compared to environmental factors (Cramer *et al.*, 1995; van Noord *et al.*, 1997). Family history is considered a good predictor of menopausal onset (Cramer *et al.*, 1995; Torgerson *et al.*, 1997; van Asselt *et al.*, 2004).

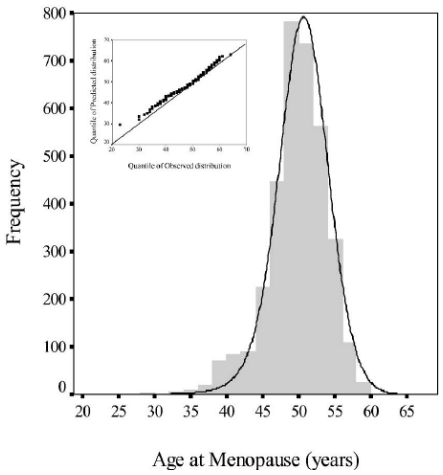


Figure 6.1: Observed frequency distribution of ages of women undergoing menopause (n=3483). Image reproduced from (Broekmans et al., 2004).

6.1.2 Oestrogen

Natural producing oestrogen, known as 17β estradiol, is a multi-disciplinary sex hormone which has several biological functions involved in endocrine and reproductive functions. In developed countries, women have longer life expectancies than men, which is believed to be a result of oestrogen (Viña *et al.*, 2005).

In the classical model of oestrogen activity, oestrogen binds to the oestrogen receptors (ERs) (typically ERα and ERβ1-6) in the cytoplasm, and then translocates to the nucleus whereby it interacts with oestrogen response elements (ERE) in specific gene DNA sequences which contain two inverted repeats. The ER complex binds AP1 enhancer elements, composed of an AP1 ligand and transcription factors, which in turn mediates gene transcription (Jensen and DeSombre, 1973; Paech *et*

al., 1997; Arnal *et al.*, 2017). Oestrogen-related receptor alpha (ERRα) is a well-characterised receptor, which leads to a series of complex downstream signalling cascades. More detail can be reviewed here (Arnal *et al.*, 2017).

6.1.3 Oestrogen and Mitochondria

Oestrogen has an important relationship with mitochondria. Mitochondria synthesise the hormone's precursor pregnenolone from cholesterol, and mitochondrial function and biogenesis is regulated by oestrogen (reviewed by (Velarde, 2014)).

Many of the mitochondrial biogenesis-regulating genes coded for by nDNA are regulated by oestrogen (Chen *et al.*, 2009). For example, oestrogen can directly upregulate transcription of nuclear-respiratory factor 1 (*NRF-1*) (Mattingly *et al.*, 2008), which is known to regulate mitochondrial biogenesis, as it mediates transcription of OXPHOS components (Jornayvaz and Shulman, 2010). NRF-1 also increases mtDNA synthesis by directly activating TFAM (Virbasius and Scarpulla, 1994; Scarpulla, 2006; Mattingly *et al.*, 2008). ERα also interacts with peroxisome proliferator-activated receptor gamma coactivator 1 (*PGC-1*), modulating expression of several of its isoforms, which is highly important in mitochondrial biogenesis regulation (Scarpulla, 2011; Kemper *et al.*, 2014). Additionally, endonuclease G (ENDOG) is directly controlled by ERRα, and has also been reported to increase mitochondrial biogenesis in cardiomyocytes, as it clusters with mitochondrial genes as well as binding mtDNA directly (McDermott-Roe *et al.*, 2011). Furthermore, oestrogen regulates mitochondrial calcium activity, which is important for mitochondrial signalling and control of cellular apoptosis through mitogen-activated protein kinase pathways (Improta-Brears *et al.*, 1999; Lobatón *et al.*, 2005).

Oestrogen has been extensively studied and has presented several therapeutic effects for protecting mitochondria. As well as upregulating the expression of antioxidant genes directly (Borrás *et al.*, 2006; Borrás *et al.*, 2010), oestrogen has demonstrated neuroprotective properties, alleviating diseases such as Alzheimer's (Nilsen *et al.*, 2006), and amyotrophic lateral sclerosis (ALS) (Choi *et al.*, 2008; Kim *et al.*, 2012), as well as cerebral vascular defects (Stirone *et al.*, 2005; Kemper *et al.*, 2014), and schizophrenia (Fink *et al.*, 1996; Kulkarni *et al.*, 2012). Targeting ER-β has also recently been presented as a therapeutic treatment to prevent mitochondrial disease LHON (Pisano *et al.*, 2015).

6.1.4 Oestrogen and Menopause

Oestrogen is predominantly synthesised by granulosa and theca cells of the ovarian follicles in women (Young and McNeilly, 2010), whereby the pregnenolone diffuses from granulosa cells in to theca cells where it is converted to androstenedione, which then diffuses back to the granulosa cells to be converted to oestrogen (Velarde, 2014).

Natural reproductive aging results in a gradual decrease in the production of oocytes from ovaries, and when follicle levels drop below approximately 1000 follicles, the menstrual cycle ceases, and women enter menopause (Velarde, 2014). Several endocrinological changes occur in the lead up to menopause, normally presenting at the age of 40 by a marked increase of follicle-stimulating hormone, which eventually results in a decrease in the sex hormones oestrogen, progesterone, and testosterone (Sherman *et al.*, 1976; Burger, 1996; Ferrini and Barrett-Connor, 1998; Stolk *et al.*, 2012).

An increased number of damaged mitochondria are reported in ovarian follicles of aged women, of which granulosa cells contain high numbers of mitochondria with ruptured membranes (Seifer *et al.*, 2002; de Bruin *et al.*, 2004). Increased mitochondrial ROS levels in mice also result in infertility, and defective folliculogenesis (Lu *et al.*, 2008). This suggests age-related decline in mitochondrial function may accelerate the decline in ovarian follicles, and contribute to reproductive ageing (Velarde, 2014).

The onset of menopause results in several age-related conditions such as bone loss (leading to osteoporosis), weight gain, hot flushes, disturbed sleep, and skin atrophy. Hormone replacement therapies (HRT) have been seen to alleviate these conditions (Velarde, 2014), however they are associated with an increased risk of developing cancer (Toniolo *et al.*, 1995; Rossouw *et al.*, 2002; Santen *et al.*, 2010).

Stolk and colleagues performed a meta-analysis of 22 genome-wide association studies to identify genes associated with menopause (Stolk *et al.*, 2012). They reported 13 novel loci, including genes responsible for DNA repair, such as mitochondrial DNA polymerase gamma (*POLG)*, as well as genes involved in immune function. Additionally, they reported biological processes linked to mitochondrial dysfunction to be related to the timing of menopause. As mtDNA CN

changes are highly dependent on POLG activity (Müller-Höcker *et al.*, 2011; Lee *et al.*, 2015), and have been reported to change with age (Cree *et al.*, 2008a; Mengel-From *et al.*, 2014), I set out to investigate changes in mtDNA CN in pre- and post-menopausal women, because this has not been studied before.

6.2 Aim

Explore the biological protective qualities of oestrogen on mitochondrial maintenance with age in UKBS females, and investigate the extent oestrogen affects mtDNA CN changes in pre- and post-menopausal women. As menopause occurs by 50 years old (te Velde *et al.*, 1998; Wallace and Kelsey, 2010), UKBS females can be stratified as pre- or post-menopausal by age, and nuclear genetic loci can be investigated to be associated with mtDNA CN changes in those groups.

6.3 Material and Methods

6.3.1 Cohort

This study only used control cohort one – UKBS data. Details of the cohort and quantification of mtDNA CN are detailed in sections 2.1 and 2.8. Average age of UKBS females was 42 years, ranging from 17 to 69 years (Figure 6.2). The average age of menopause is around 50 years old (te Velde *et al.*, 1998; Wallace and Kelsey, 2010) (Figure 6.1), therefore UKBS females were stratified into two groups - pre-menopausal if they were younger than 50 years, and post-menopausal if they were 50 years and older. Demographic analysis was conducted using IBM SPSS Statistics version 22 (IBM Corporation, 2013).

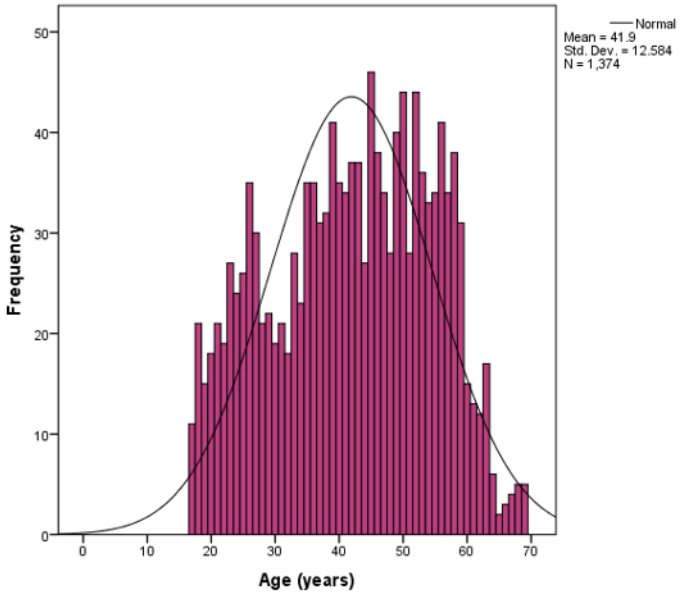


Figure 6.2: Frequency distribution of ages of UKBS females (n=1374).

6.3.2 Genotyping Preparation for Analysis

UKBS SNP genotyping data was filtered and then imputed using the Sanger Imputation Server (detailed in section 4.3.3.1.1). 2724 people (n=1358 males, n=1366 females) passed previous QC procedures, of which 2671 had logged, z-scored mtDNA CN values that were used as the phenotype.

Individual imputed chromosome VCF files returned from the Sanger Imputation Server were initially filtered to only include variants with an imputation score >=0.8 using bcftools command in SAMtools version 1.3 software (Li *et al.*, 2009). This removed 49.40% variants (Table 6.2). VCF files were then converted in to binary files using PLINK v1.90b3.44 64-bit (17 Nov 2016) (Purcell S *et al.*, 2007). Due to

formatting issues, subsequent fam files were manually updated to contain the correct family ID, sample ID, paternal ID, maternal ID, gender and phenotype (which was the logged, Z-scored mtDNA CN value) (Table 6.1).

Family ID	Sample ID	Paternal ID	Maternal ID	Gender	Phenotype
0	WTCCCT511838_74230_A08	0	0	0	-9
0	WTCCCT511842_74230_B08	0	0	0	-9
0	BLOOD292527_74236_A01	0	0	0	-9
0	BLOOD292509_74236_A02	0	0	0	-9

Family ID	Sample ID	Paternal ID	Maternal ID	Gender	Phenotype
WTCCCT511838	74230_A08	0	0	2	0.634651
WTCCCT511842	74230_B08	0	0	2	-1.09736
BLOOD292527	74236_A01	0	0	2	-0.02505
BLOOD292509	74236_A02	0	0	1	0.988776

Table 6.1: Example of binary fam files (excluding the headers); [top] incorrectly formatted, [bottom] correctly formatted. File contains family and sample IDs that were provided with UKBS samples. UKBS samples were unrelated therefore have values of zero in paternal and maternal IDs. Gender: 1 is male and 2 is female. Phenotype is logged z-scored mitochondrial DNA copy number values.

Following this, binary files were filtered for MAF <0.01, removing a further 62.55% of variants (Table 6.2). Binary files were restricted to contain only females (n=1366), removing 1358 males (49.85%). Details of the number of variants removed at each QC stage are detailed in Table 6.2.

6.3.3 Genome-wide Association Analysis

Binary files were further stratified to contain pre- (<50 years old, n=921) or post-menopausal UKBS women (>=50 years old, n=445), of which 903 and 435 had phenotype values respectively. Individual linear association tests were performed on each group using PLINK v1.90b3.44 64-bit (17 Nov 2016) (Purcell S *et al.*, 2007), where an additional filter of MAF < 0.01 was applied. In pre-menopausal women, this removed a further 0.77% of variants, whereas in post-menopausal women, this removed a further 1.43% of variants (Table 6.2). 7,373,827 variants and 7,324,188 variants passed QC in pre- and post-menopausal women respectively. Basic analysis included four principal components (PCAs), age, and panel as covariates. Full analysis included four PCAs, age, panel, absolute neutrophil, and lymphocyte counts. Output files were analysed, and Manhattan and QQ plots were generated using R version 3.1.3.

		Filtering Imputed Variants						Linear Assoc MAF < 0.01					
	Imputed Variants	*INFO > 0.08*			*MAF < 0.01*			*Pre-menopause*			*Post-menopause*		
Chr	N	Removed	Remaining	% Removed	Removed	Remaining	% Removed	Removed	Remaining	% Removed	Removed	Remaining	% Removed
1	3,071,515	1,518,764	1,552,751	49.45	984,848	567,903	63.43	4,232	563,671	0.75	8,427	559,476	1.48
2	3,393,852	1,669,052	1,724,800	49.18	1,099,693	625,107	63.76	4,999	620,108	0.80	9,379	615,728	1.50
3	2,823,295	1,355,215	1,468,080	48.00	936,972	531,108	63.82	4,169	526,939	0.78	7,324	523,784	1.38
4	2,788,969	1,368,900	1,420,069	49.08	878,847	541,222	61.89	4,202	537,020	0.78	7,829	533,393	1.45
5	2,589,304	1,237,432	1,351,872	47.79	867,164	484,708	64.15	4,098	480,610	0.85	7,409	477,299	1.53
6	2,461,245	1,160,133	1,301,112	47.14	805,009	496,103	61.87	3,522	492,581	0.71	7,681	488,422	1.55
7	2,290,419	1,113,197	1,177,222	48.60	741,968	435,254	63.03	3,475	431,779	0.80	5,723	429,531	1.31
8	2,243,785	1,087,288	1,156,497	48.46	739,193	417,304	63.92	3,383	413,921	0.81	6,128	411,176	1.47
9	1,697,071	812,643	884,428	47.89	563,888	320,540	63.76	2,571	317,969	0.80	4,885	315,655	1.52
10	1,938,280	920,556	1,017,724	47.49	637,564	380,160	62.65	2,828	377,332	0.74	5,118	375,042	1.35
11	1,950,148	948,676	1,001,472	48.65	629,769	371,703	62.88	2,707	368,996	0.73	5,697	366,006	1.53
12	1,859,555	898,932	960,623	48.34	602,199	358,424	62.69	2,996	355,428	0.84	5,011	353,413	1.40
13	1,394,148	670,409	723,739	48.09	445,778	277,961	61.59	2,098	275,863	0.75	3,613	274,348	1.30
14	1,281,139	633,517	647,622	49.45	402,309	245,313	62.12	2,184	243,129	0.89	3,856	241,457	1.57
15	1,158,409	603,105	555,304	52.06	346,585	208,719	62.41	1,464	207,255	0.70	3,004	205,715	1.44
16	1,281,946	657,773	624,173	51.31	395,901	228,272	63.43	1,881	226,391	0.82	3,320	224,952	1.45
17	1,096,020	573,603	522,417	52.34	330,066	192,351	63.18	1,439	190,912	0.75	2,889	189,462	1.50
18	1,105,391	542,358	563,033	49.06	348,350	214,683	61.87	1,621	213,062	0.76	3,333	211,350	1.55
19	868,972	467,038	401,934	53.75	238,110	163,824	59.24	1,266	162,558	0.77	2,109	161,715	1.29
20	885,404	435,657	449,747	49.20	282,674	167,073	62.85	1,215	165,858	0.73	2,073	165,000	1.24
21	531,559	268,181	263,378	50.45	160,555	102,823	60.96	775	102,048	0.75	1,448	101,375	1.41
22	524,731	267,842	256,889	51.04	155,781	101,108	60.64	711	100,397	0.70	1,219	99,889	1.21
Total	**39,235,157**	**19,210,271**	**20,024,886**	**49.40**	**12,593,223**	**7,431,663**	**62.55**	**57,836**	**7,373,827**	**0.77**	**107,475**	**7,324,188**	**1.43**

Table 6.2: Details of the number of imputed variants removed at each stage of filtering in preparation to perform genome-wide association studies on pre- and post-menopausal women using UKBS data.

6.4 Results

6.4.1 Demographic Details of Pre- and Post-Menopausal Women

1346 women (86.84%) from UKBS cohort had both SNP genotyping and mtDNA CN data (n=435 women aged 50 or older, n=911 women aged less than 50 years old).

mtDNA CN distribution between both pre- and post-menopausal women was initially assessed before starting analysis, to check for normal distribution. As Figure 6.3 shows, both raw mtDNA CN, and logged, z-scored mtDNA CN were normally distributed. To keep analysis consistent with previous work, logged, z-scored mtDNA CN was used in further analysis.

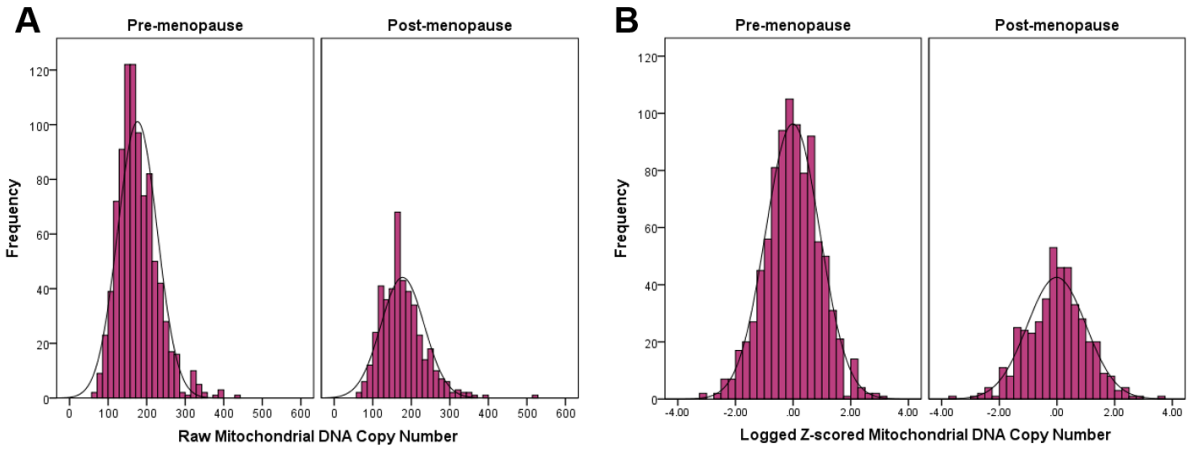


Figure 6.3: Frequency plots of [A] raw and [B] logged and Z-scored mitochondrial DNA copy number (copies/cell) distributions between pre- (age<50, n=911) and post-menopausal (age>=50, n=435) women in UKBS samples. Curves indicate normal distributions. Mitochondrial DNA copy number was calculated using the delta Ct MT-ND1: B2M calculation.

Age and mtDNA CN information in pre- and post-menopausal women is summarised in Table 6.3.

UKBS Females	Number (N)	Age Range (years)	Median Age (± standard deviation)	Mean Copy Number (copies/cell) (± standard deviation)
All	1346	17-69	43 (12.64)	176.60 (52.94)
Pre-menopause	911	17-49	36 (9.17)	176.25 (51.34)
Post-menopause	435	50-69	56 (4.44)	177.35 (56.19)

Table 6.3: Overview of ages and the distribution of raw mitochondrial DNA copy number in UKBS females with genotyping and mitochondrial DNA copy number data, stratified by pre- and post-menopause.

6.4.2 Mitochondrial DNA Copy Number Differences between Pre- and Post-Menopausal Women

To investigate any changes in mtDNA CN due to menopause, boxplots were generated to compare mtDNA CN between pre- and post-menopausal women (Figure 6.4).

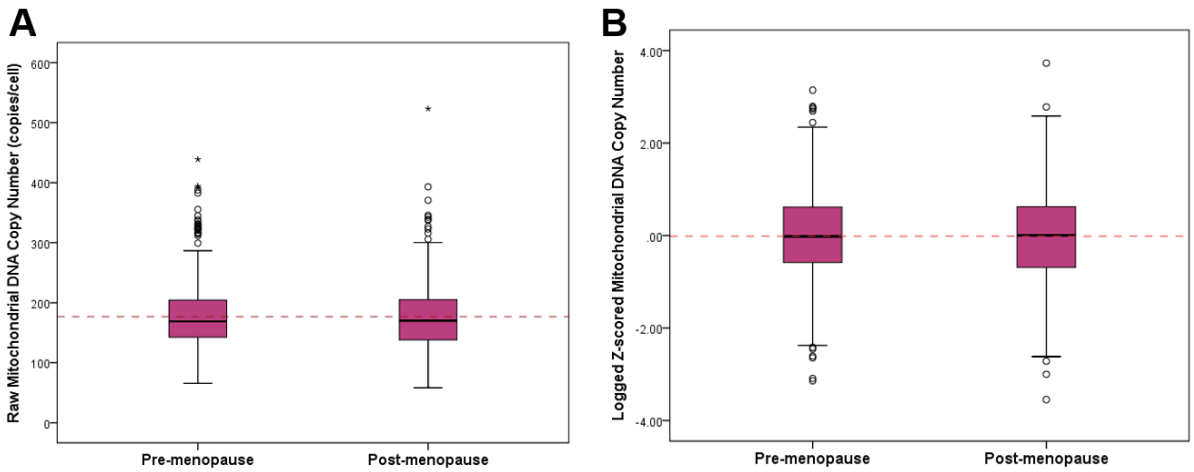
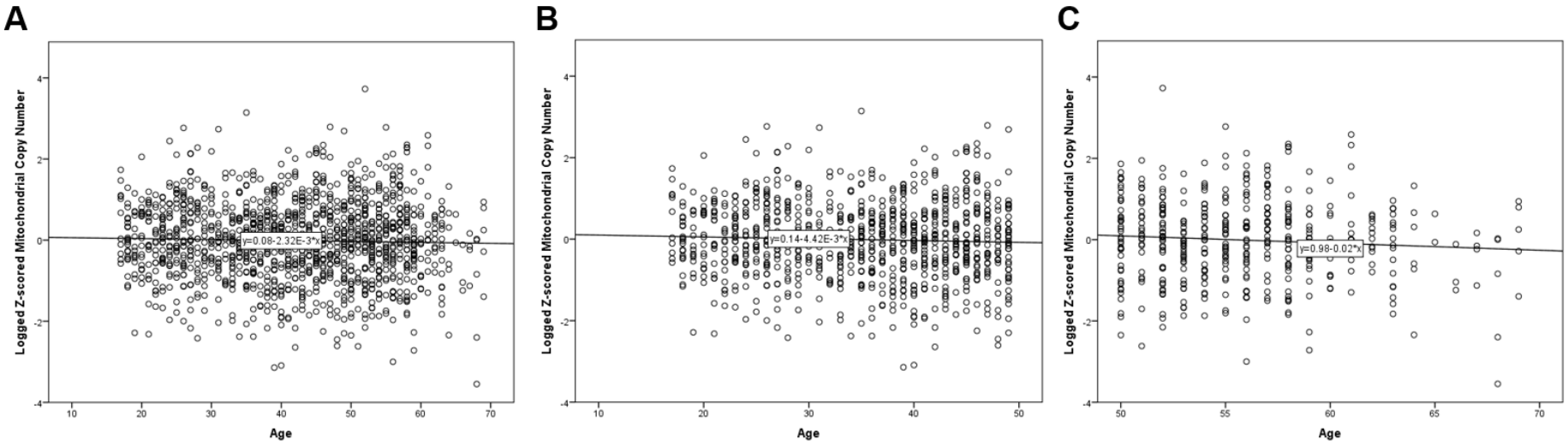


Figure 6.4: Boxplot of [A] raw and [B] logged Z-scored mitochondrial DNA copy number between pre- (age<50, n=911) and post-menopausal (age>=50, n=435) women in UKBS samples. Red dotted lines indicate mitochondrial DNA copy number population mean (μ=176.60 copies/cell).

As Figure 6.4 shows, there is no significant difference in raw mtDNA CN between pre- and post-menopausal women (independent T-test (ITT): p= 0.719, Mann-Whitney-U (MWU) test: p= 0.950).

6.4.3 Ageing Effect on Mitochondrial DNA Copy Number in Pre- and Post-Menopausal Women

To identify if there are any age-related changes in mtDNA CN between pre- and post-menopausal women, scatter plots were generated and linear regression tests were performed (Figure 6.5). As summarised by the table in Figure 6.5, there is no significant change in mtDNA CN in pre-menopausal women (r^2= 0.001, β= -0.145, *p*= 0.436), or post-menopausal women (r^2= 0.006, β= -0.962, p= 0.114). This replicates the results reported earlier in section 3.1.2.2.



UKBS Females	r^2	β	*p*
All	1.91e-04	-0.058	0.612
Pre-menopause	0.001	-0.145	0.436
Post-menopause	0.006	-0.962	0.114

Figure 6.5: Scatter plots indicating the linear relationship of logged Z-scored mitochondrial DNA copy number changes (y-axis) with age (x-axis) in UKBS females. [A] All females (n=1346). [B] Pre-menopause females (ages<50 years, n=911), and [C] post-menopause (ages<=50 years, n=435). Table below indicates respective r2, beta and p values for each linear relationship.

6.4.4 Mitochondrial Haplogroup W in Pre- and Post-Menopausal Women

In section 5.4.4, mitochondrial haplogroup W individuals had a significantly lower mtDNA CN compared to any other European mitochondrial haplogroup. To investigate whether menopause had any further impact on mtDNA CN regulation on those individuals who were mitochondrial haplogroup W, pre- and post-menopausal women were restricted to only contain haplogroup W individuals (n=13 and n=14 respectively). Boxplots were then generated to compare logged, z-scored mtDNA CN between those individuals from each group (Figure 6.6).

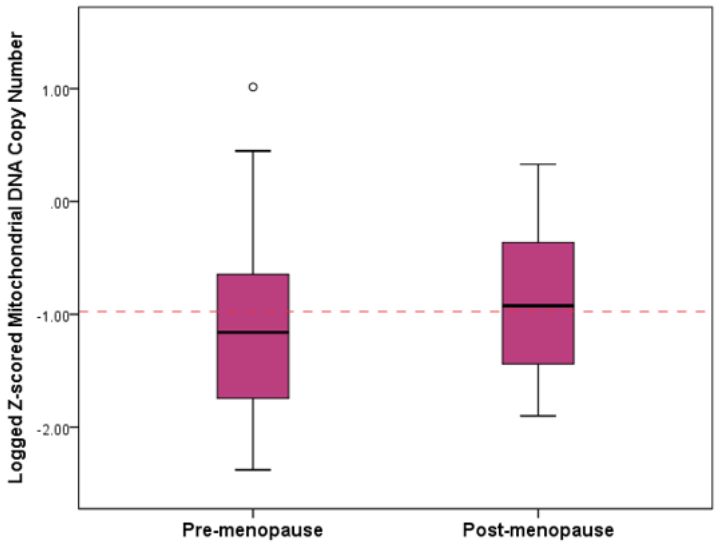


Figure 6.6: Logged Z-scored mitochondrial DNA copy number differences in mitochondrial haplogroup W pre- (n=13, μ=-1.04 ±0.97 s.d) and post-menopausal women (n=14, μ=-0.91 ±0.70 s.d). Red dotted line indicates population mean (μ=-0.97).

As Figure 6.6 shows, there is no difference in logged z-scored mtDNA CN between pre- (n=13, μ= -1.04, ±0.97 s.d) and post-menopausal (n=14, μ= -0.91, ±0.70 s.d). ITT and MWU tests also both reported that mtDNA CN was not significantly different (ITT: p= 0.699, MWU: p= 0.616).

6.4.5 Nuclear Genetic Variants Affecting Menopausal Changes

To assess nuclear genetic variants associated with mtDNA CN changes in pre- and post-menopausal women, imputed UKBS genotyping data was cleaned to only contain variants with an imputation score >=0.8, and MAF >1%. UKBS samples were stratified by pre- and post-menopausal women, and linear association analysis was performed on each group in PLINK, treating logged, z-scored mtDNA CN as a continuous phenotypic trait.

Basic analysis included four PCAs, age, and panel as covariates, whereas full analysis also included absolute neutrophil and lymphocyte counts. Manhattan and QQ plots were generated in R 3.1.3 (Figure 6.7 and Figure 6.8). Summaries of GWAS hits which had a *p* value of less than or equal to 1.00E-06 are detailed in Table 6.4 to Table 6.6. Genes were annotated to the human GRCh37-hg19 build using Ensembl (Aken *et al.*, 2016; Herrero *et al.*, 2016; Yates *et al.*, 2016).

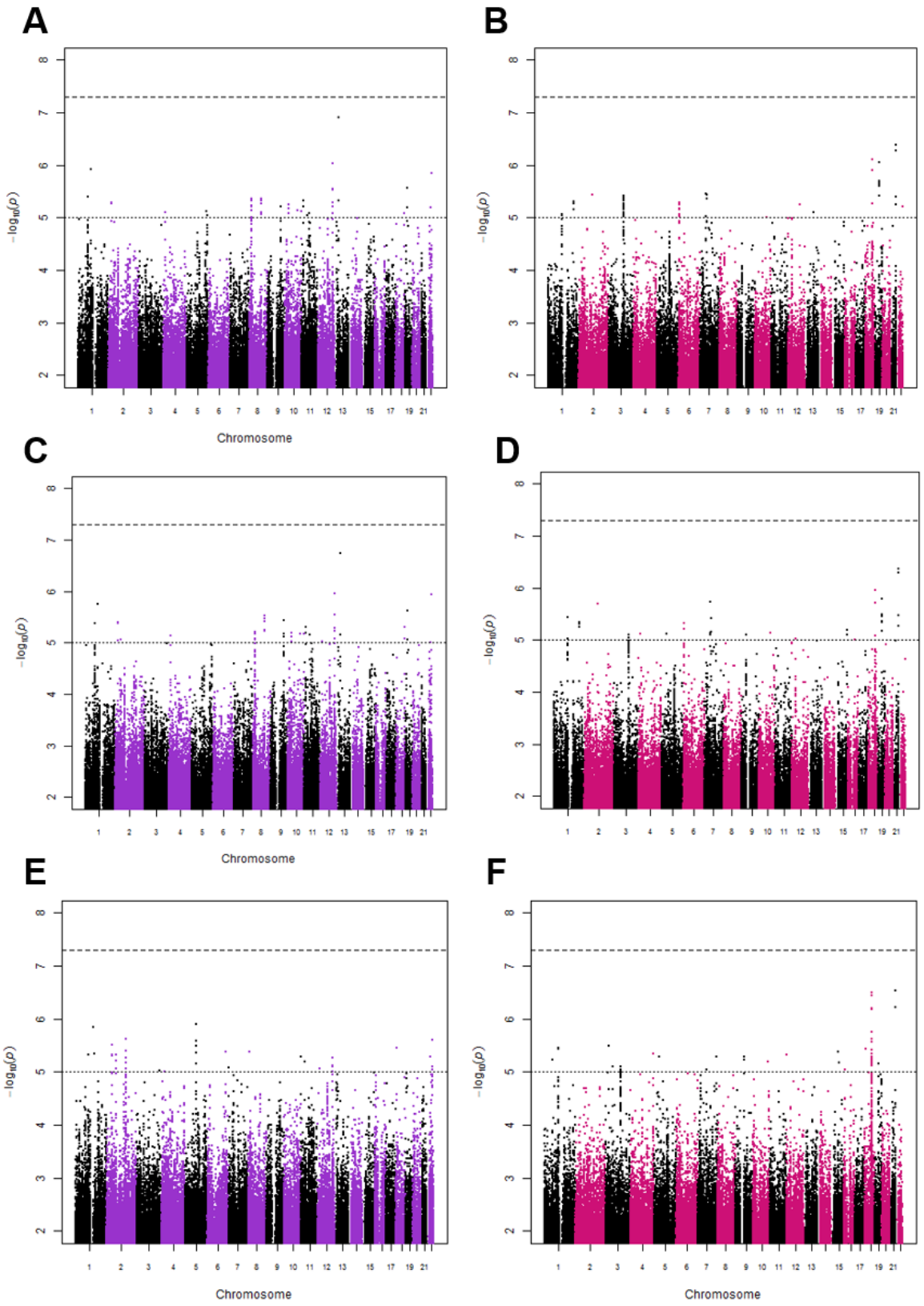


Figure 6.7: Manhattan plots of linear genome-wide association tests performed on UKBS pre- (n=903) and post-menopausal (n=435) women. [A], [C] and [E] Manhattan plots for pre-menopausal women (<50 years old). [B], [D] and [F] Manhattan plots for post-menopausal women (>= 50 years old). [A] and [B] No covariates applied. [C] and [D] Basic covariates applied including four PCAs, age and panel. [E] and [F] Full covariates applied including four PCAs, age, panel, absolute neutrophil count and absolute lymphocyte count. Respective QQ plots are in Figure 6.8. Details of top variants are in Table 6.4 and Table 6.5.

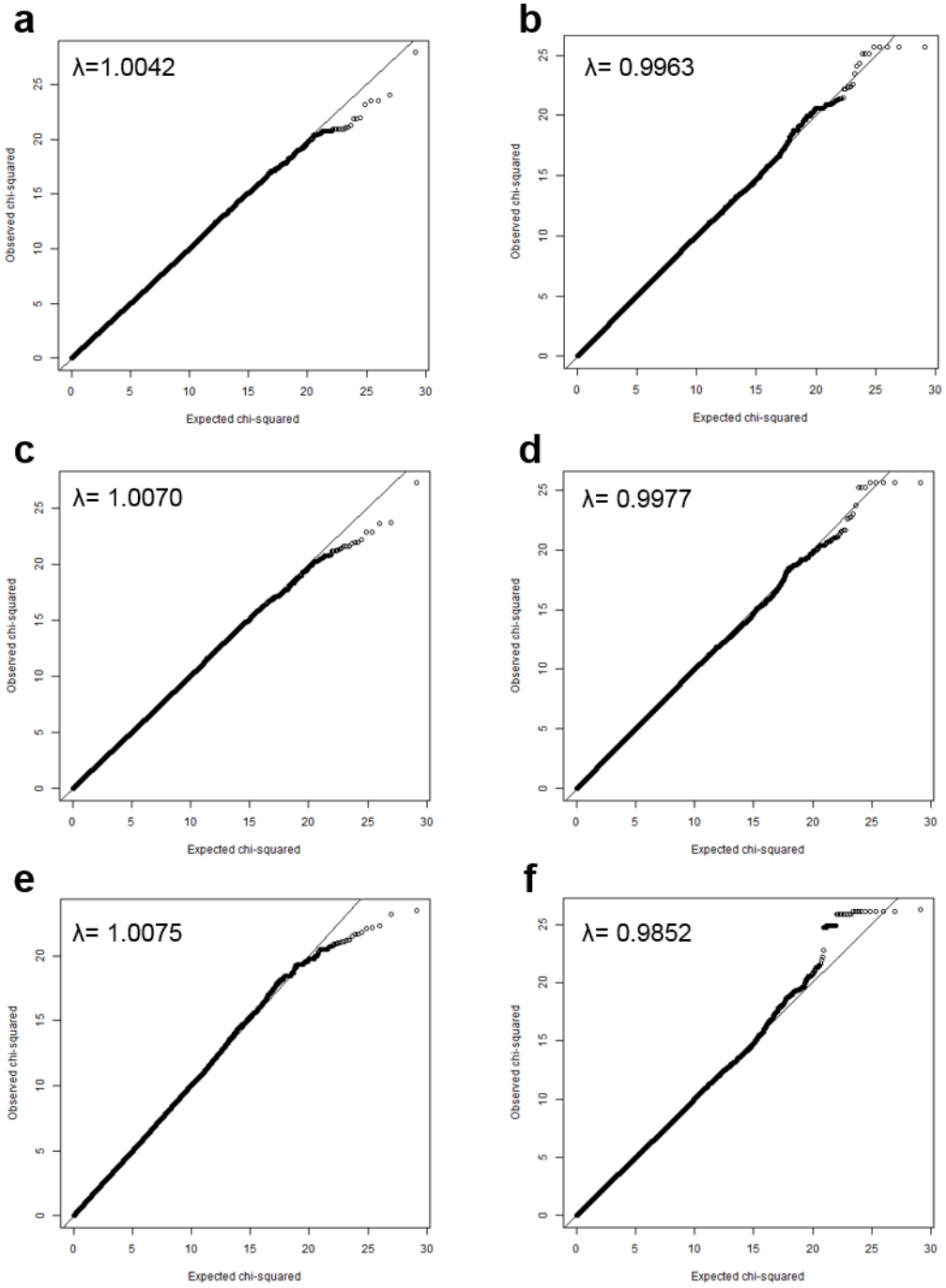


Figure 6.8: QQ plots of linear genome-wide association tests performed on UKBS pre- (n=903) and post-menopausal (n=435) women. [a], [c] and [e] QQ plots for pre-menopausal women (<50 years old). [b], [d] and [f] QQ plots for post-menopausal women (>= 50 years old). [a] and [b] No covariates applied. [c] and [d] Basic covariates applied including four PCAs, age and panel. [e] and [f] Full covariates applied including four PCAs, age, panel, absolute neutrophil count and absolute lymphocyte count. Respective Manhattan plots are in Figure 6.7. Details of top variants are in Table 6.4 and Table 6.5.

Pre-menopausal (n=903)											
CHR	**SNP**	**BP**	**A1**	**BETA**	**SE**	**L95**	**U95**	**STAT**	**P**	**Corresponding Gene**	**Neighbour Genes**
13	rs17502937	30440740	T	-0.7769	0.1458	-1.063	-0.4911	-5.328	1.25E-07	Upstream Gene Variant	*UBL3, LINC00544*
12	rs12310752	122282117	G	-0.2177	0.04404	-0.304	-0.1314	-4.943	9.17E-07	*HPD*	*PSMD9, WDR66, SETD1B, RHOF*
Post-menopausal (n=435)											
CHR	**SNP**	**BP**	**A1**	**BETA**	**SE**	**L95**	**U95**	**STAT**	**P**	**Corresponding Gene**	**Neighbour Genes**
21	rs8127035	43664085	A	0.9747	0.1894	0.6034	1.346	5.145	4.06E-07	*ABCG1*	*C21orf128, TFF3*
21	rs116494042	43664778	G	0.9747	0.1894	0.6034	1.346	5.145	4.06E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117916745	43664993	T	0.9747	0.1894	0.6034	1.346	5.145	4.06E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117798682	43665917	C	0.9747	0.1894	0.6034	1.346	5.145	4.06E-07	*ABCG1*	*C21orf128, TFF3*
21	rs116937511	43666639	G	0.9747	0.1894	0.6034	1.346	5.145	4.06E-07	*ABCG1*	*C21orf128, TFF3*
21	rs2234717	43646035	T	0.9814	0.1927	0.6038	1.359	5.094	5.25E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117364972	43656936	A	0.9814	0.1927	0.6038	1.359	5.094	5.25E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117475268	43659597	A	0.9814	0.1927	0.6038	1.359	5.094	5.25E-07	*ABCG1*	*C21orf128, TFF3*
18	rs4801111	52577579	A	-0.3673	0.07332	-0.511	-0.2236	-5.01	7.95E-07	*CCDC68*	*RAB27B*
19	rs12976863	33778706	T	0.5269	0.1057	0.3197	0.734	4.985	8.99E-07	Upstream Gene Variant	*SLC7A10, CEBPG, PEPD*

Table 6.4: Summary of top hits (p<= 1.00E-06) for genome-wide association study performed on UKBS pre- (n=903) and post-menopausal (n=435) women, where no covariates were included in the linear association model. Abbreviations - CHR: chromosome, SNP: SNP ID, BP: base position, A1: minor allele, BETA: beta value, SE: standard error, L95: lower bound of 95% confidence interval, U95: upper bound of 95% confidence interval, STAT: T-statistic and P: p-value.

Pre-menopausal (n=903)											
CHR	**SNP**	**BP**	**A1**	**BETA**	**SE**	**L95**	**U95**	**STAT**	**P**	**Corresponding Gene**	**Neighbour Genes**
13	rs17502937	30440740	T	-0.7709	0.1466	-1.058	-0.4836	-5.259	1.81E-07	Upstream Gene Variant	*UBL3, LINC00544*
Post-menopausal (n=435)											
CHR	**SNP**	**BP**	**A1**	**BETA**	**SE**	**L95**	**U95**	**STAT**	**P**	**Corresponding Gene**	**Neighbour Genes**
21	rs8127035	43664085	A	0.9769	0.1901	0.6042	1.35	5.138	4.24E-07	*ABCG1*	*C21orf128, TFF3*
21	rs116494042	43664778	G	0.9769	0.1901	0.6042	1.35	5.138	4.24E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117916745	43664993	T	0.9769	0.1901	0.6042	1.35	5.138	4.24E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117798682	43665917	C	0.9769	0.1901	0.6042	1.35	5.138	4.24E-07	*ABCG1*	*C21orf128, TFF3*
21	rs116937511	43666639	G	0.9769	0.1901	0.6042	1.35	5.138	4.24E-07	*ABCG1*	*C21orf128, TFF3*
21	rs2234717	43646035	T	0.987	0.1936	0.6076	1.366	5.099	5.15E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117364972	43656936	A	0.987	0.1936	0.6076	1.366	5.099	5.15E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117475268	43659597	A	0.987	0.1936	0.6076	1.366	5.099	5.15E-07	*ABCG1*	*C21orf128, TFF3*

Table 6.5: Summary of top hits (p<= 1.00E-06) for genome-wide association study performed on UKBS pre- (n=903) and post-menopausal (n=435) women, where basic covariates age, panel and four PCAs were included in the linear association model. Abbreviations - CHR: chromosome, SNP: SNP ID, BP: base position, A1: minor allele, BETA: beta value, SE: standard error, L95: lower bound of 95% confidence interval, U95: upper bound of 95% confidence interval, STAT: T-statistic and P: p-value.

Pre-menopausal (n=903)											
CHR	**SNP**	**BP**	**A1**	**BETA**	**SE**	**L95**	**U95**	**STAT**	**P**	**Corresponding Gene**	**Neighbour Genes**
Post-menopausal (n=435)											
CHR	**SNP**	**BP**	**A1**	**BETA**	**SE**	**L95**	**U95**	**STAT**	**P**	**Corresponding Gene**	**Neighbour Genes**
21	rs117183313	43644630	C	1.014	0.1941	0.6335	1.394	5.225	2.97E-07	*ABCG1*	*C21orf128, TFF3*
18	rs1819889	52509882	T	-0.4448	0.08541	-0.6122	-0.2773	-5.207	3.24E-07	*RAB27B*	*CCDC68*
18	rs66823280	52510205	T	-0.4448	0.08541	-0.6122	-0.2773	-5.207	3.24E-07	*RAB27B*	*CCDC68*
18	rs57603571	52511321	C	-0.4448	0.08541	-0.6122	-0.2773	-5.207	3.24E-07	*RAB27B*	*CCDC68*
18	rs62091476	52512444	C	-0.4448	0.08541	-0.6122	-0.2773	-5.207	3.24E-07	*RAB27B*	*CCDC68*
18	rs12968814	52512776	G	-0.4448	0.08541	-0.6122	-0.2773	-5.207	3.24E-07	*RAB27B*	*CCDC68*
18	rs7241733	52514226	C	-0.4448	0.08541	-0.6122	-0.2773	-5.207	3.24E-07	*RAB27B*	*CCDC68*
18	rs7242037	52514408	A	-0.4448	0.08541	-0.6122	-0.2773	-5.207	3.24E-07	*RAB27B*	*CCDC68*
18	rs1978355	52515885	A	-0.4448	0.08541	-0.6122	-0.2773	-5.207	3.24E-07	*RAB27B*	*CCDC68*
18	rs11152021	52515974	C	-0.4448	0.08541	-0.6122	-0.2773	-5.207	3.24E-07	*RAB27B*	*CCDC68*
18	rs2016584	52485773	A	-0.4477	0.08635	-0.6169	-0.2784	-5.185	3.62E-07	*RAB27B*	*CCDC68*
18	rs2006985	52503100	G	-0.4477	0.08635	-0.6169	-0.2784	-5.185	3.62E-07	*RAB27B*	*CCDC68*
18	rs1972592	52505326	A	-0.4477	0.08635	-0.6169	-0.2784	-5.185	3.62E-07	*RAB27B*	*CCDC68*
18	rs1972593	52505867	A	-0.4477	0.08635	-0.6169	-0.2784	-5.185	3.62E-07	*RAB27B*	*CCDC68*
18	rs1972594	52505882	A	-0.4477	0.08635	-0.6169	-0.2784	-5.185	3.62E-07	*RAB27B*	*CCDC68*
18	rs7241381	52505962	A	-0.4477	0.08635	-0.6169	-0.2784	-5.185	3.62E-07	*RAB27B*	*CCDC68*
18	rs1972595	52506077	A	-0.4477	0.08635	-0.6169	-0.2784	-5.185	3.62E-07	*RAB27B*	*CCDC68*
18	rs12607380	52507078	T	-0.4477	0.08635	-0.6169	-0.2784	-5.185	3.62E-07	*RAB27B*	*CCDC68*
18	.	52507270	A	-0.4477	0.08635	-0.6169	-0.2784	-5.185	3.62E-07	*RAB27B*	*CCDC68*
18	rs1833285	52507877	T	-0.4477	0.08635	-0.6169	-0.2784	-5.185	3.62E-07	*RAB27B*	*CCDC68*
21	rs2234717	43646035	T	1.006	0.1979	0.6184	1.394	5.084	5.95E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117364972	43656936	A	1.006	0.1979	0.6184	1.394	5.084	5.95E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117475268	43659597	A	1.006	0.1979	0.6184	1.394	5.084	5.95E-07	*ABCG1*	*C21orf128, TFF3*
21	rs8127035	43664085	A	0.9865	0.194	0.6062	1.367	5.084	5.95E-07	*ABCG1*	*C21orf128, TFF3*
21	rs116494042	43664778	G	0.9865	0.194	0.6062	1.367	5.084	5.95E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117916745	43664993	T	0.9865	0.194	0.6062	1.367	5.084	5.95E-07	*ABCG1*	*C21orf128, TFF3*
21	rs117798682	43665917	C	0.9865	0.194	0.6062	1.367	5.084	5.95E-07	*ABCG1*	*C21orf128, TFF3*
21	rs116937511	43666639	G	0.9865	0.194	0.6062	1.367	5.084	5.95E-07	*ABCG1*	*C21orf128, TFF3*
21	rs11909889	43667961	A	0.9865	0.194	0.6062	1.367	5.084	5.95E-07	*ABCG1*	*C21orf128, TFF3*
21	rs71320516	43670845	C	0.9865	0.194	0.6062	1.367	5.084	5.95E-07	*ABCG1*	*C21orf128, TFF3*
21	rs116850768	43670906	T	0.9865	0.194	0.6062	1.367	5.084	5.95E-07	*ABCG1*	*C21orf128, TFF3*
18	rs2008739	52483302	A	-0.4424	0.08729	-0.6135	-0.2713	-5.068	6.45E-07	*RAB27B*	*CCDC68*
18	rs118123068	51851415	T	-0.8992	0.1776	-1.247	-0.5511	-5.063	6.62E-07	*STARD6*	*POLI, C18orf54*
18	rs75605963	51907649	T	-0.8992	0.1776	-1.247	-0.5511	-5.063	6.62E-07	*STARD6*	*POLI, C18orf54*
18	rs2311120	52477123	G	-0.4419	0.0873	-0.613	-0.2708	-5.062	6.64E-07	*RAB27B*	*CCDC68*

Table 6.6: Summary of top hits (p<= 1.00E-06) for genome-wide association study performed on UKBS pre- (n=903) and post-menopausal (n=435) women, where full covariates age, panel, four PCAs and absolute neutrophil and lymphocyte counts were included in the linear association model. Abbreviations - CHR: chromosome, SNP: SNP ID, BP: base position, A1: minor allele, BETA: beta value, SE: standard error, L95: lower bound of 95% confidence interval, U95: upper bound of 95% confidence interval, STAT: T-statistic and P: p-value.

As Table 6.5 shows, in pre-menopausal women, an upstream gene variant on chromosome 13 which neighbours *UBL3* and *LINC00544* is a consistent top hit (p<1.00E-06) when adjusted for PCAs, age, and panel. This hit is lost when absolute neutrophil and lymphocyte counts are included as covariates (Table 6.6). Table 6.5 and Table 6.6 consistently report top hits (p <=1.00E-06) in post-menopausal women in *ABCG1* and *RAB27B,* with the addition of *STARD6* in neutrophil and lymphocyte adjusted analysis (Table 6.6).

6.5 Discussion

This is the first study which has investigated mtDNA CN changes in healthy women who are either pre- or post-menopause. Despite the role of oestrogen in mitochondrial biogenesis regulation (section 6.1.3), my results highlight that there is no significant difference between mtDNA CN in pre- or post-menopausal women. There is also no significant change in mtDNA CN with age in pre- and post-menopausal women, and supports my previous findings in section 3.3.4, which report no mtDNA CN changes with age in females, but a statistically significant, albeit likely not biologically significant, decrease in males with age. Women who were defined as mitochondrial haplogroup W, and were previously reported to have a significantly lower mtDNA CN (section 5.4.4), are shown here to have a similar mtDNA CN in both pre- and post-menopausal women. This suggests that mtDNA variants are not contributing to menopausal changes. A potential explanation for this may be due to medications which may or may not have been provided to individuals during menopause. Unfortunately these details were not available to us.

Despite using a large well-defined female cohort, there are several limitations which may explain a negative result. Although we stratified our samples by the average age of menopause (te Velde *et al.*, 1998; Wallace and Kelsey, 2010), menopause is actually stratified into three stages – pre-, peri- and post-menopause, and it is dependent on menstrual bleeding patterns and hormone levels. Oestrogen levels are variable between women as they are in the peri-menopausal stage. This introduces another category that should have been stratified by in this study to identify more accurate differences in mtDNA CN changes, rather than just using age as a factor to categorise individuals (Sherman *et al.*, 1976). Stage and age of onset of menopause would need to be provided to make this study more precise.

HRTs and contraceptives have also been reported to alter oestrogen levels, and affect mtDNA CN changes in women. Knez and colleagues reported a significantly higher mtDNA CN in females than males, but reported a significantly lower mtDNA CN in women taking systemic hormone therapies, including contraceptives, and post-menopausal therapies (Knez *et al.*, 2016). Therapeutics may explain why I record no difference in mtDNA CN, because if both groups have females taking hormonal medication, the average mtDNA CN will be lowered to a central mean, showing no differences. Medication details would need to be provided to enable analysis to be correctly stratified, presuming the medication effected mtDNA CN changes.

The effect of ageing on mtDNA CN has been well documented, and changes are believed to be a result of prolonged ROS exposure causing genomic insults (Cree *et al.*, 2008a; Hegde *et al.*, 2012; Mengel-From *et al.*, 2014). Oestrogen binds to ERs on the surface of cells, which activates MAP kinases, which activate nuclear factor kappa B (*NF-κB*), which in turn binds to the promotors of superoxide dismutase (SOD) and glutathione peroxidase (GPx), and upregulates their levels. SOD and GPx neutralise damaging hydrogen peroxide (H_2O_2) compounds, and significantly higher levels of SOD and GPx are recorded in female rats (Viña *et al.*, 2005). Viña and colleagues also reported significantly lower levels of mtDNA damage in female Wistar rats compared to males, and the increased SOD and GPx levels, stimulated by oestrogen production, may be a contributing factor for decreased mtDNA damage over time. This may explain why no change in mtDNA CN is reported in UKBS females, because if less mtDNA damage has accumulated in females over time whilst oestrogen was being produced, and SOD and GPx were being generated as a result, insufficient time between pre- and post-menopause had passed to allow the effects of ROS to impact mtDNA CN changes. Further work would need to be pursued to investigate this further.

It would have been informative to analyse mtDNA CN in blood cell components between pre- and post-menopausal women. Bain and colleagues report higher WBC in African women who take oral contraceptives compared to those not taking oral contraceptives (Bain *et al.*, 1984). Osteoporosis is a common problem in post-menopausal women, who are not taking HRT, however oestrogen therapies alleviate this (Burger, 1996; Ebeling *et al.*, 1996). Miyazaki argue that mtDNA CN is important for maintaining overall bone health, as it oversees osteoclast survival (Miyazaki *et al.*,

2012). This suggests a close relationship between oestrogen, blood cell components, and mtDNA CN.

In the GWAS analysis conducted between pre- and post-menopausal women, only post-menopausal women had results that were consistent across full and basic analysis. In basic adjusted analysis, a strong hit (p <=1.00E-06) in *ABCG1* was reported, with positive beta values, suggesting a positive control on mtDNA CN (Table 6.5). *ABCG1* is a member of the ATP binding cassette family, which are a group of proteins responsible for transporting various molecules across intra and extra-membranes. *ABCG1* is a member of the white subfamily, which is involved in macrophage cholesterol and phospholipid trafficking (NCBI Gene ID: 9619 details).

When absolute neutrophil and lymphocyte counts were included as covariates in the GWAS model, adjusting for cell count, *RAB27B* and *STARD6* were also included as loci associated with mtDNA CN changes in post-menopausal women (Table 6.6), however they both had a negative correlation with mtDNA CN. Rab proteins regulate vesicular transport steps between membrane-bounded organelles, and are thought to be important in targeting, docking, and/or fusion of vesicular carriers with their specific acceptor compartments (Simons and Zerial, 1993; Pfeffer, 2001). Rab proteins are cell specific, and can act as systemic clocks, regulating exchange of molecules like nucleotides between compartments (Zerial and McBride, 2001). Certain Rabs have been identified to control mitochondrial relocation by lysosomal trafficking (Jin and Mills, 2014), and loss of *RAB27B* results in damaged mitochondria (Chiang *et al.*, 2011).

STARD6 is domain 6 of the STeroidogenic Acute Regulatory- (StAR) related lipid transfer family. This family are highly conserved between plants and animals, and are involved in regulating non-vesicular trafficking of lipids and sterols (Létourneau *et al.*, 2016). *STARD6* specifically binds testosterone, and mutations in StARs results in absent steroidogenesis, which causes several physiological defects including XY sex reversal and enlarged adrenals (Miller, 2010).

Although the literature reports no direct links between these genes and mitochondrial biogenesis affecting mtDNA CN, several functional pathways could have downstream effects. As previously described, oestrogen and sex hormones are synthesised by cholesterol and heavily rely on mitochondria to synthesise their precursors (Velarde,

2014). Enhanced mitochondrial respiration was reported in rats in oestrous stage compared to ovex rats (Gigli and Bussmann, 2001), and brain tissue from oestrogen-treated ovex rats also showed increased mitochondrial respiration (Nilsen *et al.*, 2006). This suggests that oestrogen increases mitochondrial respiration, and may contribute to the steady mtDNA CN previously reported.

During the initial synthesis of sex hormones, importing cholesterol from the OMM to the IMM in mitochondria is a rate-limiting step (Miller, 2013; Velarde, 2014). Acute regulation of steroidogenesis is managed by StARs, whereas chronic regulation is managed by cytochrome P450 enzymes present on both the mitochondria and ER. Supporting this, the conversion of cholesterol to pregnenolone is catalysed by cytochrome P450, and involves the oxidation of NADPH (Miller and Auchus, 2011). Lewis and colleagues previously reported that interactions between the ER and mitochondria were thought to be regulating mtDNA synthesis (Lewis *et al.*, 2016), however this might be an observation of oestrogen metabolism effects on the mitochondria.

Overexpression of StARs decreases intracellular lipids and lowers inflammatory factors (Ning *et al.*, 2009). During menopause, oestrogen levels decrease, which could lead to accumulation of cholesterol and other lipids in mitochondria. This could result in increased mitochondrial stress and ROS production, which could contribute to defective vesicle trafficking, and could stimulate immune responses. This might explain the negative beta association values of post-menopausal mitochondrial biogenesis with vesicle trafficking proteins *STARD6* and *RAB27B.*

Sex hormones stimulate the hypothalamic-pituitary-adrenal gland (HPA), modulating stress responses through corticosteroid release (Wilder, 1995; Whitacre, 2001). This would imply that in post-menopausal women mitochondrial respiration would be decreased, which could further lead to the accumulation of ROS. Malik and colleagues propose that mtDNA CN is regulated by ROS levels in mitochondria (Malik and Czajka, 2013), which may suggest why post-menopausal mitochondrial biogenesis is associated with these genes.

Furthermore, Graham suggests that changes in expression of mitochondrial fusion proteins are the limiting factors to cholesterol efflux efficiency, whereby defects in fusion machinery results in inflammation and neurological disorders (Graham, 2015).

The decrease of oestrogen production may subject the mitochondria to further physiological stresses, and the inability of mitochondria to adapt, and fuse together may contribute to the overall stress signals generated by the cell.

Hormone receptors are on the surface of immune cells (Da Silva, 1995), but cytokine receptors IL-1R and IL-18R are also on the surface of hormone-producing tissue, which suggests there is a feedback system between the two (Fairweather and Rose, 2004). We report a lower mtDNA CN as WBC count increases, which is supported by other studies (Knez *et al.*, 2016). I previously reported differences in mtDNA CN relative to the composition of WBCs in blood (section 3.3.5 - Figure 3.14 and Table 3.4), which play a role in immune responses. Following the study published by Stolk and colleagues, who reported immune genes to be involved in menopause onset, further work should investigate the relationship between mtDNA CN changes and oestrogen levels, and immune markers to see if there is any correlation (Stolk *et al.*, 2012).

Despite the plausible biological explanation of *RAB27B*, *ABCG1,* and *STARD6* being associated with post-menopausal women in this study, Stolk and colleagues do not report any of these genes in their meta-analysis study of menopause across 22 GWAS (Stolk *et al.*, 2012). This study only uses hundreds of individuals to perform GWAS analysis, which is statistically underpowered to identify genes regulating a complex trait like mtDNA CN. Larger sample sizes would be needed for future work.

Additionally, a cohort with more elderly women (> 69 years) would need to be used to identify real age and oestrogen-related effects on mtDNA CN after menopause. This study identified women as pre- or post-menopausal using age as a predictor, however this means this study is only 75% accurate because 75% of women undergo menopause before 50 years (Broekmans *et al.*, 2004). Ideally I could have stratified the cohort as women under 60 years old to be considered pre-menopausal, however stratifying at higher ages would have significantly reduced the number of samples for comparison, as UKBS females have a 17-69 years age range (Figure 6.2), therefore this was not possible.

This study would also need to be replicated in an independent, control cohort. Ideally I would have used the ALSPAC cohort as a replicate cohort, however all of the

females in this cohort are of child bearing age (16-43 years old), and had not undergone menopause yet, and was therefore incomparable.

Furthermore, mtDNA CN is reported to significantly decrease in females with age in the literature (Lee *et al.*, 2010; Mengel-From *et al.*, 2014; Knez *et al.*, 2016). This study opposes these results though, as I report no significant change in mtDNA CN in women with age. This poses the question – are the results reported a real indication of oestrogen or age effects?

In conclusion, this is the first study to investigate mtDNA CN changes between pre- and post-menopausal women. From this work, I have identified limitations that future studies need to consider, whilst potentially identifying viable genes that could be contributing to menopausal changes in women in relation to mitochondrial biogenesis regulation.

Chapter 7 Validating Mitochondrial DNA Deletions

7.1 Introduction

The mutation rate in mtDNA is 10-20 times higher than that in nDNA, which causes point mutations and large-scale deletions to sporadically develop (Richter et al., 1988; Lynch et al., 2006). 13-30% of all cases of mitochondrial disease, result from mtDNA deletions (Krishnan *et al.*, 2007; Grady *et al.*, 2014a), however heteroplasmy levels are a contributing factor (Chinnery and Hudson, 2013). Mitochondrial deletion-related diseases manifest from lower heteroplasmic thresholds (approximately 50–60%) than other mitochondrial diseases (Rossignol et al., 2003).

Mitochondrial deletions present as single deletions or as multiple deletions (Krishnan et al., 2008), and deletion breakpoints have been mapped (Damas et al., 2014). Single deletions are rare, and present in 1-20 people out of every 100,000 in the general public (Chen et al., 1995), and give rise to sporadic mitochondrial disease such as progressive external opthalmoplegia, Kearns Sayre syndrome, and Pearson's syndrome (Chinnery et al., 2004). Large-scale deletions are more common in mtDNA than small-scale deletions because it is believed that smaller mtDNA molecules can replicate, and accumulate, faster (Diaz et al., 2002).

It has been reported that almost half of human oocytes contain 5-10 deleted molecules per 100,000 molecules of mtDNA, but only one in 8,000 individuals end up harbouring pathogenic deletions during development because of the mtDNA CN bottleneck effect (Chinnery et al., 2004). Generally, multiple point mutation mitochondrial deletions accumulate in post-mitotic tissues, and high energy demanding tissues such as muscle, with age because mtDNA are exposed to genetic insults from factors like ROS (Sciacco *et al.*, 1994; Bender *et al.*, 2006; Kraytsberg *et al.*, 2006). This, combined with the lack of recombination in mtDNA, also poses a disadvantage as mitochondrial deletions accumulate deleterious mutations in a 'Muller's Ratchet', whereby germline selection is the only method to counteract their accumulation (Muller, 1964; Aanen *et al.*, 2014). This is assessed by cell fitness, and the ratio of wild-type to mutant mtDNA determines the overall biochemical defects (Stewart and Chinnery, 2015).

The mechanisms by which mitochondrial deletions develop is still unclear, however interruptions in mtDNA replication machinery contribute to mitochondrial deletion development. For example, decreases in nucleotide pooling, and defects in

replication machinery like polymerase-gamma (POLG) result in mtDNA deletion (Saada, 2004; Pitceathly et al., 2012; Nikkanen et al., 2016).

Mitochondrial deletions vary between tissue type (Fischel-Ghodsian *et al.*, 1992), and contribute to heteroplasmy levels, although they do not always result in disease (Chinnery *et al.*, 2004). This is believed to be a result of clonal expansion regulation, where accumulation of mutant mtDNA results in defective OXPHOS. In post-mitotic cells such as neurons, the combined accumulation of mtDNA mutations and loss of cells from ageing results in neurodegenerative diseases, such as Alzheimer's disease and Parkinson's disease (Bender *et al.*, 2006). mtDNA deletions have also been seen to repopulate quicker in post-mitotic cells because as metabolic activity decreases from loss of wildtype mtDNA, causing increased replication of mutant mtDNA and slower replication of wildtype mtDNA, resulting in an accumulation of mutant DNA, increased ROS, and ageing (Diaz *et al.*, 2002). As a result, overall mtDNA CN increases (Chen *et al.*, 2011).

mtDNA CN, but not mtDNA point mutations, have been argued to be responsible for neurodegenerative disease development (Keeney and Bennett, 2010; Giordano *et al.*, 2014; Rice *et al.*, 2014; Pyle *et al.*, 2015a; Wei *et al.*, 2017), and a recent study also reported specific mutant mtDNA (harbouring a 3.1kb deletion) in *C. elegans* (known as *uaDf5)* replicated itself by exploiting wildtype mtDNA machinery, despite *uaDf5* being destructive to the cell. They report the mitochondrial unfolded protein response is responsible for *uaDf5* levels, as well as mtDNA CN control (Gitschlag *et al.*, 2016). Overall providing evidence that mtDNA deletions, and mtDNA CN, have a mutual relationship.

In addition to studying mtDNA CN, I took the opportunity to assess the levels of mitochondrial deletions in a large population sample, and to investigate mtDNA deletions in relation to mtDNA CN changes, using DNA extracted from white blood cells. Despite a study reporting that peripheral blood contains approximately 75% abnormal mtDNA molecules (Fischel-Ghodsian *et al.*, 1992), most studies have reported blood from healthy individuals to contain no, or few, mitochondrial deletions (Maagaard *et al.*, 2006; Van Haute *et al.*, 2013). This is likely to be because of the short lifespan of white blood cells, as they are renewed every 1-5 days (Flindt, 2006), allowing little time for mitochondrial deletions to accumulate. To date, no study has investigated baseline mitochondrial deletion levels in a large population control

cohort. As my DNA samples were all extracted from white blood cell pellets from healthy donors, and based on the published literature, it would be unexpected to detect any mitochondrial deletions in these samples. Therefore I decided to explore this.

The qPCR method used to quantify mtDNA CN (section 2.8.1.4), was originally designed to additionally measure mtDNA deletions by measuring the proportion of *MT-ND1* in the minor arc to *MT-ND4* in the major arc (He *et al.*, 2002; Krishnan *et al.*, 2007; Grady *et al.*, 2014a; Rygiel *et al.*, 2015). Approximately 85% of mitochondrial deletions sit within the major arc region of mtDNA, located between the origin of light strand (O_L) and heavy strand (O_H) replication (Shoffner *et al.*, 1989; Payne *et al.*, 2011; Phillips *et al.*, 2014; Rygiel *et al.*, 2015). Moreover, approximately 96% of all deletions found in patients with COX deficiency were located in the *MT-ND4* region (He *et al.*, 2002). Therefore, by utilizing this method, calculating mitochondrial deletion in each sample, I could also validate the reliability of the qPCR assay whilst investigating the prevalence of mitochondrial deletions in European blood samples.

7.2 Aim

Identify if mtDNA extracted from white blood cell pellets from the blood of healthy individuals in a large, European cohort contain mitochondrial deletions. Long-range PCR (LR-PCR) will be used to visualise mtDNA deletion abundance, and validate the accuracy of established qPCR methods to calculate mitochondrial deletion levels.

7.3 Materials and Methods

This study was conducted using cohort one - UKBS (n=3091). Full details of the cohort and mtDNA CN calculations are in Chapter 2. There was insufficient DNA available to analyse cohort two, and the qPCR technique used to determine mtDNA CN in the cohort three - ALSPAC cohort used a single mtDNA probe and was therefore unable to detect mitochondrial deletions (section 2.8.3).

7.3.1 Identifying UKBS Samples that Potentially Contain Mitochondrial Deletions

mtDNA CN values were calculated using either *MT-ND1* or *MT-ND4* as the mitochondrial reference gene. As detailed in section 3.3.2 - Figure 3.5, UKBS samples had their mtDNA CN determined by the delta Ct calculation using either; the mean mitochondrial reference gene Ct values adjusted to the mean housekeeping gene (*B2M*) Ct value, or using the raw median Ct of the mitochondrial reference gene adjusted to the raw median Ct of the housekeeping gene (*B2M*).

As indicated in section 3.3.2, *MT-ND1* derived values were plotted against respective *MT-ND4* values to both validate the qPCR assay, and identify if any samples contained mitochondrial deletions (Figure 7.1). Linear regression lines were plotted to identify correlation, along with 95% confidence, and prediction intervals. Samples that sat outside of the prediction intervals indicated they contained a deletion (~20%). Samples which sat to the left of the lines (n=185) indicated a deletion in *MT-ND1*, and those to the right (n=61) indicated a deletion in *MT-ND4* (Figure 7.1A). However this was not replicated when looking at the raw median Ct data of these samples (Figure 7.1B), suggesting that an artificial deletion effect may have been introduced as a result of the qPCR QC steps, and adjustment to *B2M*. To validate these ‘deletions’, further tests needed to be performed. As most mitochondrial deletions sit within the *MT-ND4* region, I decided to validate the samples that sat to the right of the line (n=61).

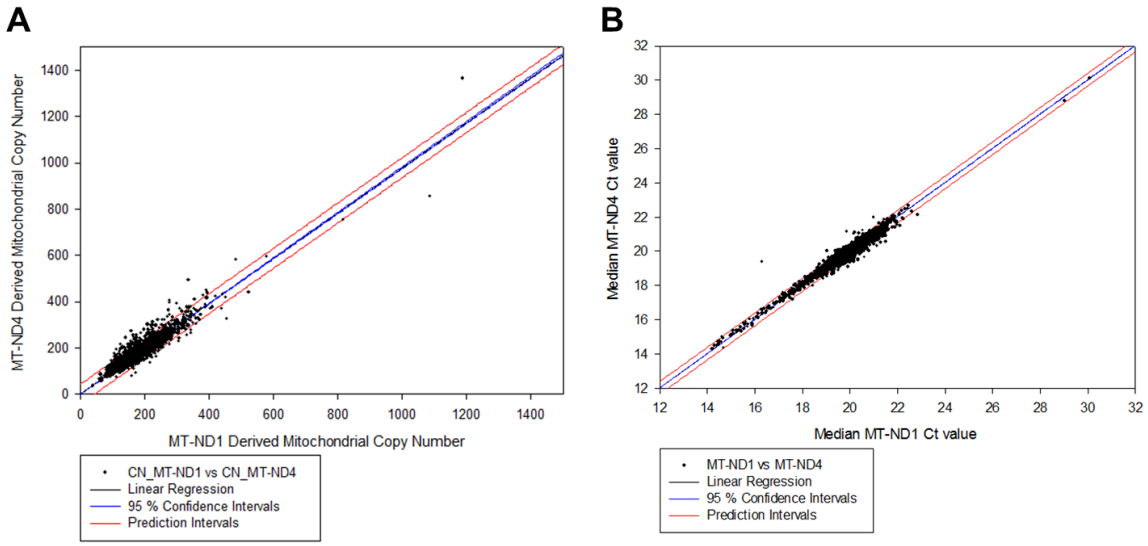


Figure 7.1: Linear regression plot of [A] Mitochondrial DNA copy number derived from using mitochondrial gene MT-ND1 against mitochondrial DNA copy number derived from using mitochondrial gene MT-ND4 in UKBS samples (n=3028), and [B] raw median Ct values generated from triplicate measurements for MT-ND1 plotted against raw median MT-ND4 measurements for each UKBS sample (n=3088).

7.3.2 Detecting Mitochondrial Deletions using Quantitative-PCR

Standard protocols quantify mitochondrial deletions by measuring the proportion of *MT-ND4* to *MT-ND1* in a multiplex qPCR reaction (He *et al.*, 2002; Krishnan *et al.*, 2007; Phillips *et al.*, 2014; Rygiel *et al.*, 2015). As detailed in the section 2.8.1.4, *B2M*, *MT-ND1,* and *MT-ND4* were measured in a multiplex qPCR reaction in all UKBS samples. Following QC procedures, where thresholds and standards were set and replicates greater than 0.5 Ct removed; mitochondrial deletions were quantified for each sample using Equation 7.1. Table 7.1 gives an example of the analysis performed on each sample across each qPCR run to calculate each sample's mitochondrial deletion percentage (deln (ddCt)). If the ratio of *MT-ND4* to *MT-ND1* is 0.9 (90%), this signifies that 10% of mtDNA molecules measured may contain a mitochondrial deletion in *MT-ND4* (Rygiel et al., 2015).

$$ddCt = MT\text{-}ND4\ Ct\ Mean\ [(Control) - (Sample)] - MT\text{-}ND1\ Ct\ Mean\ [(Control) - (Sample)]$$

$$deln = 1 - (2^{ddCt})$$

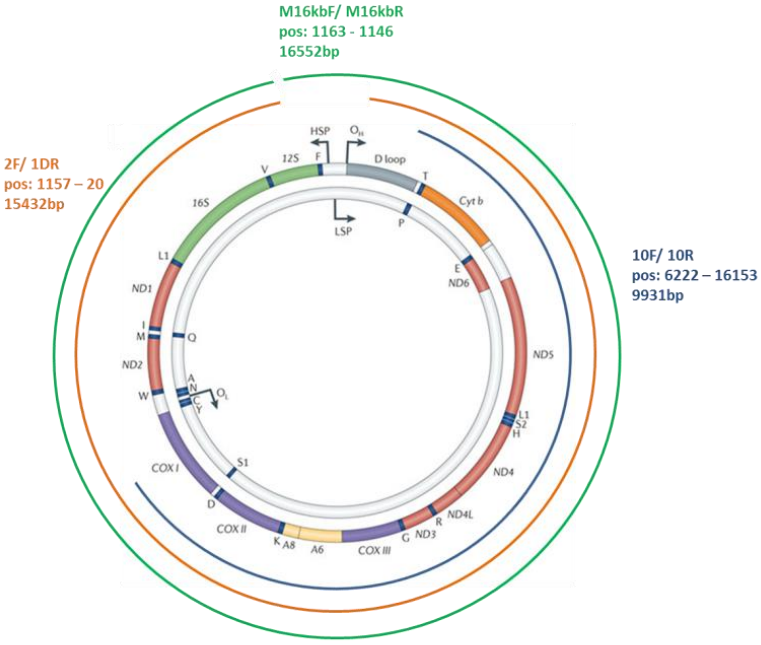
Equation 7.1: Equation used to calculate mitochondrial deletions. ddCt is the difference between the MT-ND4 mean Ct between the control and the sample and the MT-ND1 mean Ct between the control and the sample. Overall deletion is one minus two to the power of ddCt.

		MTND1							MTND4								
Run_ID	ID	Raw_Ct	Ct_Mean	Ct_St_Dev	SQ	Log_SQ	SQ_Mean	SQ_St_Dev	Raw_Ct	Ct_Mean	Ct_St_Dev	SQ	Log_SQ	SQ_Mean	SQ_St_Dev	ddCt	deln (ddCt)
Run 1	Control (1/100)	23.607	23.701	0.088	51185.889	4.709	48209.382	2779.742	23.658	23.710	0.051	51786.063	4.714	50115.577	1640.323		
Run 1	Control (1/100)	23.714	23.701	0.088	47761.448	4.679	48209.382	2779.742	23.711	23.710	0.051	50053.487	4.699	50115.577	1640.323		
Run 1	Control (1/100)	23.782	23.701	0.088	45680.807	4.660	48209.382	2779.742	23.761	23.710	0.051	48507.180	4.686	50115.577	1640.323		
Run 1	Deletion (1/100)	18.004	17.938	0.093	1952079.630	6.290	2039272.352	123309.130	19.529	19.447	0.117	719333.098	5.857	759162.764	56327.654		
Run 1	Deletion (1/100)	17.872	17.938	0.093	2126465.074	6.328	2039272.352	123309.130	19.364	19.447	0.117	798992.430	5.903	759162.764	56327.654	-1.500	0.646
Run 1	Sample_1	19.860	20.066	0.178	584179.804	5.767	513390.737	61305.945	19.538	19.639	0.090	715464.832	5.855	671286.116	39188.414		
Run 1	Sample_1	20.170	20.066	0.178	477680.234	5.679	513390.737	61305.945	19.711	19.639	0.090	640716.629	5.807	671286.116	39188.414		
Run 1	Sample_1	20.168	20.066	0.178	478312.174	5.680	513390.737	61305.945	19.670	19.639	0.090	657676.887	5.818	671286.116	39188.414	0.436	-0.353
Run 1	Sample 2	20.205	20.278	0.083	467030.966	5.669	445897.204	23769.912	19.958	19.988	0.035	547285.624	5.738	537100.632	12026.360		
Run 1	Sample 2	20.368	20.278	0.083	420163.525	5.623	445897.204	23769.912	20.027	19.988	0.035	523833.004	5.719	537100.632	12026.360		
Run 1	Sample 2	20.260	20.278	0.083	450497.121	5.654	445897.204	23769.912	19.979	19.988	0.035	540183.269	5.733	537100.632	12026.360	0.299	-0.230

Table 7.1: An example of the table used to calculate mitochondrial deletion percentages for each UKBS sample. Raw data for each sample from qPCR analysis was imported in to the table for MT-ND1-HEX and MT-ND4-Cy5 and calculations were applied respectively (Equation 7.1). ddCt is the difference between the MT-ND4 mean Ct between the control and the sample and the MT-ND1 mean Ct between the control and the sample. Overall deletion is one minus two to the power of ddCt and is a decimal percentage. Note – the deletion sample only has two replicates because one replicate was removed during QC. Analysis was performed in Microsoft Excel 2010.

7.3.3 Visualising and Verifying Mitochondrial Deletions

LR-PCR was performed on UKBS samples which had a mitochondrial deletion percentage >20%, because these samples were outside the mtDNA CN population confidence intervals, suggesting they contained mitochondrial deletions (He *et al.*, 2002) (Figure 7.1). LR-PCR was performed across two different mitochondrial regions spanning the major arc - mitochondrial deletion amplicon one (ref: 10R/10F) of 9931 bp (Figure 7.2 blue line), and amplicon two (ref: 2F/D1R) of 15432 bp (Figure 7.2 orange line) using primers detailed in Figure 7.2 table.



Long-Range Reference	Primer Binding Sites	Amplicon Size (bp)	Forward Primer Sequence (5′-3′)	Reverse Primer Sequence (5′-3′)
10F/ 10R	6222-6240 and 16133-16153	9931	CCCTCTCTCCTACTCCTG	CAGGTGGTCAAGTATTTATGG
2F/ D1R	1-20 and 1157-1177	15432	TTAAAACTCAAAGGACCTGGC	AGGGTGATAGACCTGTGATC

Figure 7.2: Top: A schematic representation of three different sized long-range PCR amplicon regions in the mitochondrial DNA used to detect any mitochondrial deletions. The table details 5'-3' forward and reverse primer sequences used to amplify both the 10F/10R and 2F/D1R regions within mitochondrial DNA.

7.3.3.1 Optimising LR-PCR assay to DNA extracted from blood

Similar LR-PCR methods have been described previously (Payne *et al.*, 2011), however initial assay optimisation was needed. Therefore a selection of control and deletion DNA extracted from different tissues (see Figure 7.3) were used to identify optimal conditions to perform LR-PCR across different tissues, especially blood.

7.3.3.1.1 Sample Sources

As deletion assay controls, seven samples were used representing confirmed non-deleted and deleted samples. d5d25, d5AM017.7, motor neuron, and cerebellum samples were kindly donated by Dr Marzena Kurzawa-Akanbi from Newcastle University, UK in collaboration with Dr George Tofaris from University of Oxford, UK. d5d25 and d5AM017.7 DNA samples were extracted from differentiated dopaminergic neuronal cell lines which had been programmed in to induced pluripotent stem cells (iPSCs) from donor control fibroblasts using specialised iPSC technology (developed by Dr George Tofaris at University of Oxford). DNA was extracted from cell lines 25 days after differentiation from iPSCs (in vitro). Motor neuron DNA was extracted from motor cortex tissue from a donor who had Alzheimer's disease and was known to have a mitochondrial deletion. Cerebellum DNA was extracted from cerebellum tissue from a healthy donor and did not have any mitochondrial deletions detected. WT_546081 was a UKBS sample, and control calibre and deletion calibre were DNA as previously detailed in section 2.8.1.4. WT_546081 was used to determine optimum assay conditions for all future tests and calibre DNA was used to assess specificity (see Figure 7.3).

Sample ID	Control/ Deletion	mtDNA Source
WT_546081	Control	Blood
d5d25	Control	iPSC cell line
d5AM017.7	Control	iPSC cell line
Motor Neuron	Deletion	Brain tissue
Cerebellum	Control	Brain tissue
Control (calibre)	Control	Blood
Deletion (calibre)	Deletion	Cybrid cell lines

Figure 7.3: A table summarising the source of and the type (control or deletion) of mtDNA used to optimise the LR-PCR assay. iPSC cell line, motor neurons and cerebellum tissue extracted DNA were kindly provided by Dr Marzena Kurzawa-Akanbi (Newcastle University, UK) and Dr George Tofaris (University of Oxford, UK).

7.3.3.1.2 Long Range PCR Reactions

Each DNA sample was diluted to 40 ng/µl in nuclease free water (Thermo Fisher Scientific, MA, USA) and kept at -20°C until use. Fresh mastermix was made for each reaction batch, and 10F/10R and D1R/2F reactions were always run simultaneously for each sample. All reactions were set up on ice. Mastermix consisted of 1x PS PrimeSTAR GXL buffer (contains 1x Mg^{2+}), 0.625 units PrimeSTAR GXL DNA polymerase, 0.2 µM of dNTP mix (all Takara Clontech, USA), and 0.2 µM of each forward and reverse primer (Integrated DNA Technologies in Leuven, Belgium). Mastermix was aliquoted into plastic strip tubes (Starlab, Milton Keynes, UK) and approximately 2 ng of each respective DNA sample was added to each tube, and sealed securely. Samples were spun down before being loaded into an Applied Biosystems® Veriti® 96 well thermal cycler (Life Technologies, Paisley, UK). LR-PCR cycling conditions are detailed in Figure 7.4.

7.3.3.1.3 Long Range PCR Agarose Gel Electrophoresis

Following amplification of samples, 12 µl of each LR-PCR product was combined with 6 µl orange G loading buffer (Orange G powder, 50% v/v glycerol (both Sigma Aldrich Company Ltd, Dorset, UK) and 50% v/v water). A DNA negative LR-PCR reaction was also mixed with orange G loading buffer, as was a 1 Kb Plus DNA ladder (Promega, Southampton, UK). Each UKBS sample LR-PCR product was loaded in to individual wells on a 16 well comb on a large, 0.7% agarose gel (0.7% w/v agarose (Bioline, London, UK) in 1x tris-acetate-EDTA (TAE) buffer (5 Prime GmbH, Hilden, Germany) and 0.4 mg/µl UltraPure™ ethidium bromide (Invitrogen, Paisley, UK)). 1 kb Plus DNA ladder (Promega, Southampton, UK) and negative products were loaded separately. Agarose gels were initially electrophoresed at 50 V for 15 minutes and then at 40 V for three hours in 1x TAE buffer. Agarose gels were imaged and analysed using the UVP GelDoc-It™ imaging system and UVP Launch Vision Works LS software (UVP, Cambridge, UK).

9931bp amplicon (10R/10F)		
Temperature (°C)	Time	
94	1 minute	
98	10 seconds	30 cycles
60	15 seconds	
68	10 minutes*	
72	10 minutes	
4	∞	

15432bp amplicon (D1R/2F)		
Temperature (°C)	Time	
94	1 minute	
98	10 seconds	30 cycles
68	10 minutes	
72	10 minutes	
4	∞	

*Figure 7.4: LR-PCR PCR reaction conditions for each specified amplicon. * indicates change from 13 minutes to 10 minutes (Figure 7.5A and B).*

As Figure 7.5 shows, preliminary tests were performed on different tissue samples (as detailed in Figure 7.3) to identify optimal LR-PCR conditions. Figure 7.5A displays an unknown band of approximately 2000 bp present in all samples which have been amplified using D1R/2F primers, and an unknown band of approximately 2500 bp is present in all samples amplified using 10F/10R primers. These bands were thought to be generated from the annealing time being too long, therefore LR-PCR was repeated on the same tissues but the annealing temperature was reduced from 13 minutes to 10 minutes for D1R/2F reactions (Figure 7.5B). As a result, the original unspecific bands did diminish slightly, however this might be hard to interpret from Figure 7.5B as the image is overexposed.

As Figure 7.5B shows, there were still some remnants of unspecific bands in all tissues LR-PCR products, therefore a final repeat was performed on the same tissues, but each primer concentration was decreased to 0.12 µM for all reactions, instead of 0.20 µM. The annealing temperature remained at 10 minutes for D1R/2F. As Figure 7.5C shows, there were no more unspecific bands in all the LR-PCR products.

Following discussion about these results, it was decided that further LR-PCR would use PCR conditions described in Figure 7.4, and primer concentrations of 0.20 µM because DNA from blood samples amplified well under those conditions.

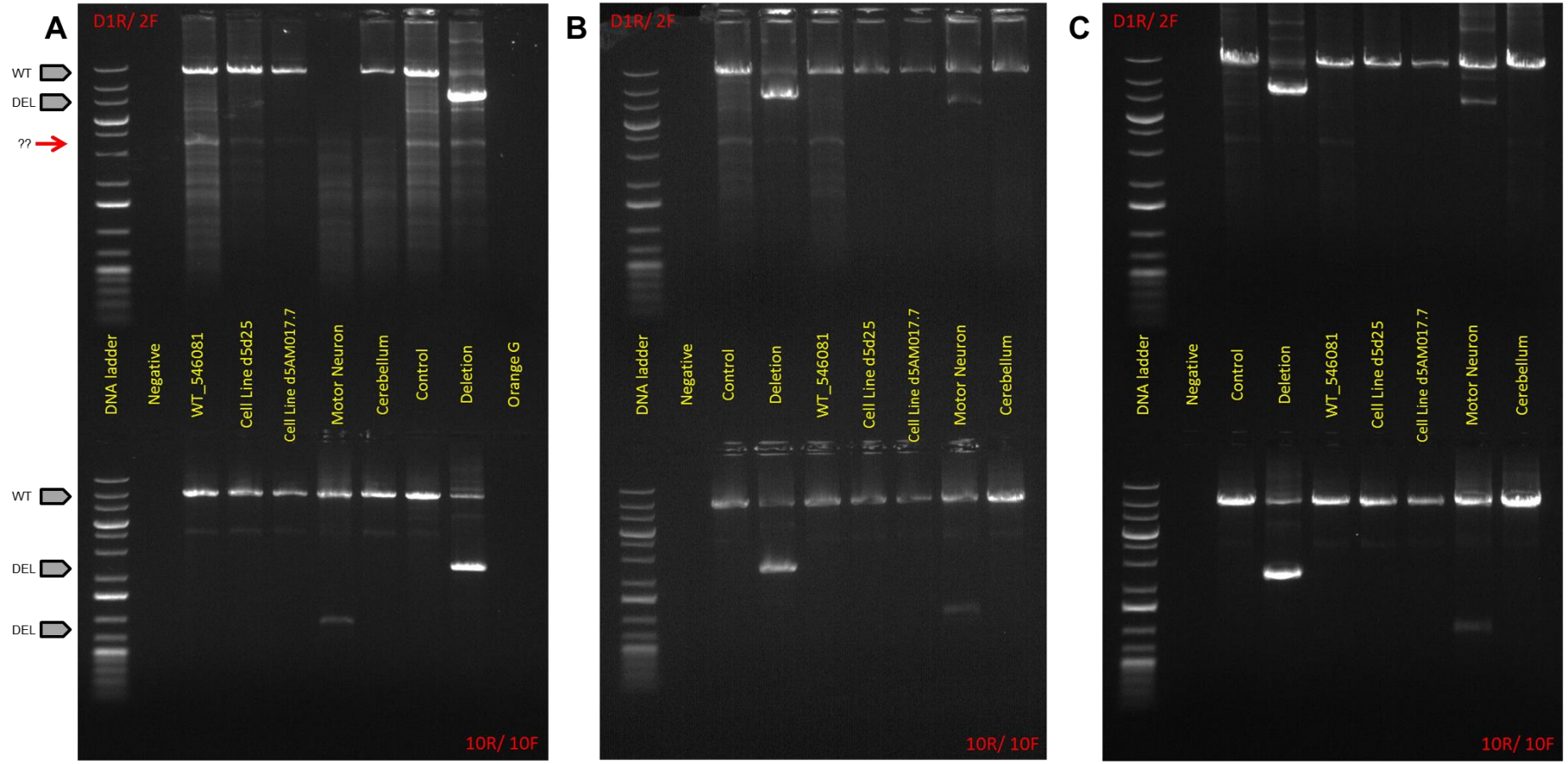


Figure 7.5: Images of LR-PCR products during optimisation steps when performed on a selection of samples. Top images show products using D1R/2F primers. Bottom images show products using 10F/10R primers. [A] images LR-PCR products when D1R/2F had an annealing time of 13 minutes at 68°C. [B] and [C] images LR-PCR products when D1R/2F had an annealing time of 10 minutes at 68°C. All primer concentrations in [A] and [B] were 0.2µM whereas all primer concentrations in [C] were 0.12 µM. WT= wildtype. DEL= deletion. Red arrow indicates an unspecific band (approximately 3500 bp) during optimisation steps. WT_546081 is a UKBS sample from blood. d5d25 and d5AM017.7 are dopaminergic neuron cell lines differentiated from iPSCs. Motor and cerebellum samples were extracted from human brain tissue. Control DNA is from healthy blood and deletion calibre is from cybrid cell lines (section 2.8.1.4). Images were generated on a UVP GelDoc-It™ imaging system and UVP Launch Vision Works LS software (UVP, Cambridge, UK).

7.4 Results

Following protocol optimisation, LR-PCR was performed on 61 UKBS samples (n=39 males and n=22 females) (approximately 2% of all samples), which had a mitochondrial deletion percentage of greater than 20%, which was calculated from qPCR tests (Table 7.2). Samples with these results were suspected to contain a mtDNA deletion in the *MT-ND4* gene located in the major arc region.

7.4.1 Detecting Mitochondrial Deletions in UKBS samples

Each UKBS sample was defrosted from -20°C, vortex and spun down and then freshly diluted to 40 ng/µl in nuclease free water before undergoing LR-PCR. As described previously, all reactions were set up fresh, on ice, and both D1R/2F and 10F/10R reactions were prepared and run at the same time on each respective sample to prevent delay leading to sample degradation. LR-PCR reactions were run overnight and kept at 4°C until the morning where they were then vortexed and spun down before being prepared for gel electrophoresis (as detailed above). D1R/2F products were run from the top of each gel, and 10F/10R products were run from the centre of each gel.

Table 7.2 lists all the UKBS samples which underwent LR-PCR. Average age in these samples was 40 (range from 15-65 years). The largest deletion reported using qPCR analysis was 52.65%.

Figure 7.6 shows UV images of the DNA fragments generated when performing LR-PCR on each UKBS sample. No results reported any deletions in any of the samples, in either of the amplicon regions. The unspecific band appeared again in two LR-PCR runs for D1R/2F (Figure 7.6A and D), however they were not indications of mtDNA deletions because the control calibre also contained the band.

WTCCC_ID	Gender	Age	mtDNA deletion size (%)
547201	M	40	20.14
546181	F	30	21.51
540337	M	50	24.80
549190	F	55	24.69
540349	F	40	21.76
541129	M	55	23.36
543044	M	45	46.84
605219	M	25	52.65
544091	F	55	24.31
604175	M	20	22.19
546081	F	40	30.70
546098	M	15	28.73
548339	F	35	20.81
548254	F	35	20.77
605176	M	35	28.75
543333	M	55	36.61
548218	M	65	22.80
547350	M	35	23.97
547272	F	15	25.15
604169	F	25	37.44
542019	F	55	21.74
544255	M	15	26.42
540050	M	45	20.86
547010	F	55	20.80
547144	F	15	30.68
547278	F	25	23.47
544182	M	55	20.33
541057	M	35	20.54
548077	M	55	22.71
548255	F	25	20.14
542191	M	35	23.63
541135	M	50	24.61
543206	M	45	23.80
548261	M	65	33.53
549068	F	45	23.07
605113	M	50	24.79
542306	M	35	23.72
542106	M	55	25.10
547327	M	65	28.10
547051	M	35	25.22
541133	F	35	24.78
541055	F	40	25.82
549079	F	40	24.34
547043	M	60	21.82
605236	F	30	22.48
547037	M	60	27.14
544002	F	40	20.49
604053	M	55	24.71
542184	M	35	22.76
549020	F	25	21.05
542161	M	50	20.86
547246	M	35	20.56
543161	M	45	26.53
544290	M	35	24.35
541466	M	45	21.80
546197	M	30	26.58
544323	M	35	24.45
544318	F	40	22.98
604195	M	45	23.35
547061	M	35	21.41
547224	M	25	25.99

Table 7.2: A list of all UKBS samples (n=61) which had a mitochondrial deletion percentage calculated to be greater than 20% using qPCR (Equation 7.1). The table lists the UKBS sample Wellcome Trust Case Control Consortium identifier (WTCCC_ID), their gender (M=male, F= female), age and mitochondrial deletion size.

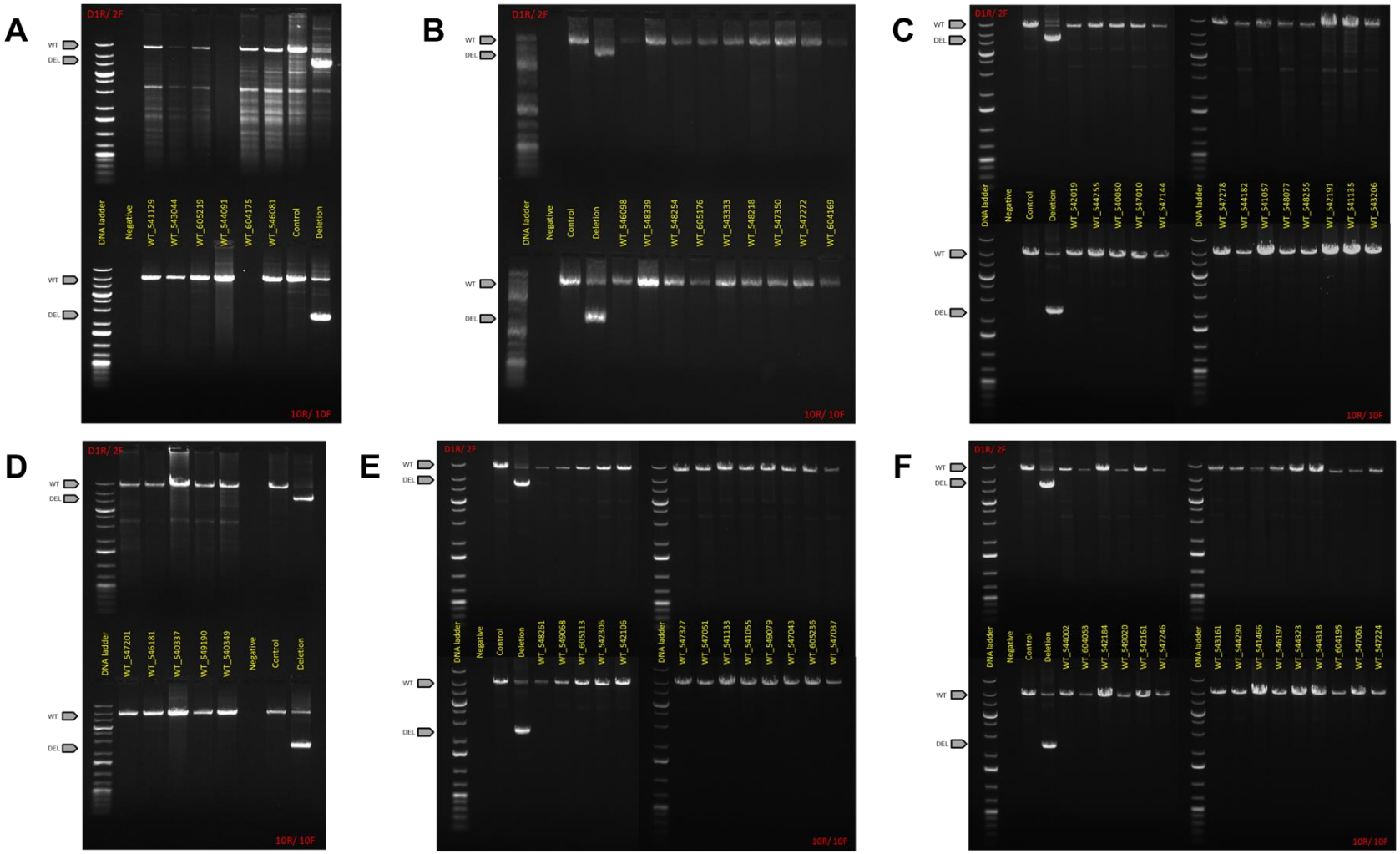


Figure 7.6: Images of UKBS samples (n=61) where long-range PCR has been used to identify deletions within the mitochondrial DNA. The top half of each image are products which have been amplified using D1R/2F primers (15432 bp region) and the bottom half of each image are products which have been amplified by 10F/10R primers (9931 bp region). Arrows indicate the wildtype (WT) and deletion (DEL) fragments. [B]-[E] are separate gel images but all were performed under the same conditions. Only [A] was performed with a longer annealing time of 13 minutes instead of 10 minutes for D1R/2F.

Validating Mitochondrial DNA Deletions

7.5 Discussion

This is the first study to investigate the abundance of mtDNA deletions in healthy individuals from a large, European population. My results show that DNA extracted from white blood cell pellets from healthy individuals contain no mitochondrial deletions, which opposes the deletion levels determined using established qPCR calculations.

Previous studies have reported DNA samples extracted from peripheral blood to contain no mitochondrial deletions (Van Haute *et al.*, 2013), which suggests that the results reported in the LR-PCR analysis are accurate, because no deletion fragments are recorded. However, this introduces limitations about using qPCR to detect mitochondrial deletions because I report deletion values up to 52.65% using qPCR, which could not be detected using LR-PCR.

qPCR is a very sensitive method, and a number of factors can affect the accuracy and reproducibility of its results (reviewed in section 4.5.6.2.1). In theory, *MT-ND1* should be proportional to *MT-ND4* in complete, circular mtDNA, therefore a decrease in *MT-ND1* would result in a decrease of *MT-ND4* relatively. To calculate deletions in *MT-ND4* in healthy individuals using qPCR, which do not have any mitochondrial deletions, would suggest there was either a higher proportion of *MT-ND1* in the cells than *MT-ND4*, or there were errors with *MT-ND4* amplification properties. If the former was true, then it may suggest there are linear molecules of mtDNA present in DNA samples, which are offsetting the proportions between *MT-ND1* to *MT-ND4* ratios. Linear molecules of mtDNA could easily be generated from DNA handling processes, and freeze-thaw damage (reviewed in section 4.5.6.2.2).

Previous work conducted in section 5.4.6 identified the importance of correct qPCR primer, and probe design for optimal, and reproducible mtDNA analysis. Expanding on this work, I wanted to investigate how the detection of mitochondrial deletions changed when using different mtDNA reference genes. I used the ‘alternative’ qPCR method (detailed in section 5.3.6.1) to quantify mitochondrial deletions in the same 108 UKBS samples (n=53 haplogroup W and n=55 non-haplogroup W) used for original analysis. Mitochondrial deletions were quantified by calculating the proportion of *MT-RNR2* to *MT-ND4*, and samples were stratified by haplogroup W to ensure mitochondrial reference genes were not having an effect on deletion levels (as mtDNA CN was previously reported to be significantly lower in haplogroup W

individuals – section 5.4.4). Mitochondrial deletion values were calculated using both qPCR methods, and then compared (Figure 7.7).

As Figure 7.7A shows, mitochondrial deletion values are supposedly significantly high (p= 2.026E-12) in haplogroup W individuals using the *MT-ND1*: *MT-ND4* calculation, however the significant difference is not reported (p= 0.457) when using the *MT-RNR2*: *MT-ND4* calculation (Figure 7.7B). These results would support the work reported using LR-PCR (Figure 7.6). Figure 7.7C also shows there is a much larger spread of deletions using the *MT-ND1*: *MT-ND4* calculations compared to the *MT-RNR2*: *MT-ND4* calculations.

These data demonstrate that incorrect mitochondrial deletion values can be reported as a result of technical limitations, and emphasises the importance of optimising qPCR assays before attempting to validate mtDNA characteristics. Furthermore, polymorphisms are reported across several *MT-ND* genes (Mitchell *et al.*, 2006), and this could potentially lead to inefficient quantification of amplicon regions in qPCR which could result in false-positive deletion results. The evidence shown in this study suggests the established mitochondrial deletion assay needs to be revised, and alternative mtDNA reference gene regions need to be identified before quantifying exact mitochondrial deletion levels.

Alternative methods being used to validate mitochondrial deletion levels include Sanger sequencing, deep sequencing, and digital-droplet PCR (ddPCR) (Ameur *et al.*, 2011; Van Haute *et al.*, 2013; Belmonte *et al.*, 2016; Rygiel *et al.*, 2016). These techniques are considered to be more accurate than qPCR because both Sanger sequencing and deep sequencing specifically sequence the mtDNA to identify genetic breakpoints and rearrangement regions. ddPCR also uses fluorescent probes and droplet technology to quantify individual mtDNA molecules to accurately determine deletions. Furthermore, Taylor and colleagues developed a method which quantified mitochondrial deletions by sequencing mtDNA in single-cells. They used 28 pairs of primers to generate overlapping fragment reads which spanned the entire mtDNA. From this, they were able to map exact mitochondrial deletion breakpoints, and regions of rearrangements (Taylor et al., 2001).

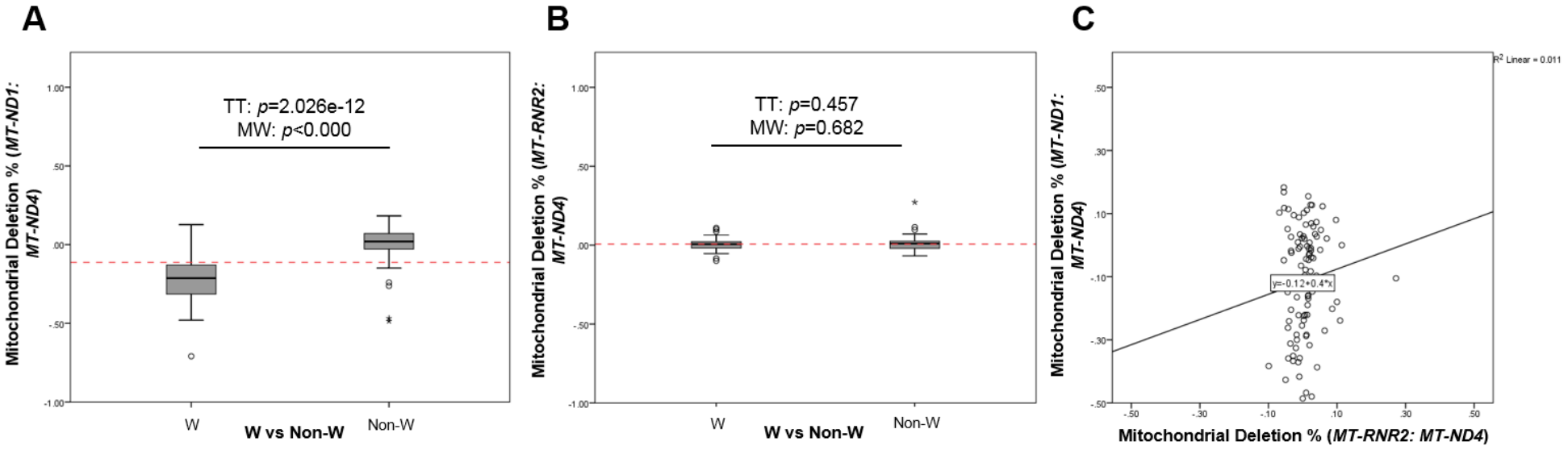


Figure 7.7: Mitochondrial deletion percentage calculated using qPCR methods which either quantified [A] MT-ND1: MT-ND4 or [B] MT-RNR2: MT-ND4 in UKBS samples (n=108). [C] Correlation between mitochondrial deletion levels using different qPCR methods.

Moreover, multiple mitochondrial regions should be used to validate mitochondrial deletion levels. Rygiel and colleagues developed a qPCR assay in 2015 which measures three mtDNA regions to validate mitochondrial deletions across a number of different tissue types (Rygiel et al., 2015), however they use the D-loop as a reference region which may cause some inconsistencies because of its associations to quadruplex mtDNA stranding issues (Takamatsu et al., 2002). Ideally a qPCR assay needs to be designed which quantifies more than three mtDNA reference regions that are not located in any highly genetic or structurally polymorphic regions.

Despite using Takara PrimeSTAR GXL reagents to perform LR-PCR on DNA samples, which is considered to be the best polymerase for long-range amplification (Jia *et al.*, 2014), an unspecific 4 kbp band was consistently reported in Figure 7.5, and in Figure 7.6A and D. It would be interesting to expand on these results, and investigate mitochondrial deletions in those UKBS samples with extra LR-PCR bands, but by using more accurate techniques like deep sequencing or ddPCR to be more comprehensive.

In summary, this study has identified that there are no mitochondrial deletions in DNA of white blood cells in general, healthy, European populations. As a result of this work, I have identified the limitations of current methods used to determine mitochondrial deletions, and propose that developments are needed to further improve the accuracy of mitochondrial deletion detection using qPCR.

Chapter 8 General Discussion

mtDNA CN is central to both cellular and mitochondrial homeostasis, as mtDNA composition dictates OXPHOS efficiency, and is vital for cellular bioenergetics. mtDNA CN varies between cell and tissue type (Michaels *et al.*, 1982; Robin and Wong, 1988; Kelly *et al.*, 2012), and its regulation has been the focus of much research, as it is used as an indicator of disease susceptibility, progression, and outcome (Cree *et al.*, 2008a; Pyle *et al.*, 2010; Malik *et al.*, 2011; Giordano *et al.*, 2014; Reznik *et al.*, 2016). Additionally, circulating, cell-free (ccf)-mtDNA has more recently received increased attention as a biomarker for disease (Yu, 2012; Liu *et al.*, 2015a; Pyle *et al.*, 2015a; Pyle *et al.*, 2015b; Li *et al.*, 2016), thus emphasising the importance of understanding mtDNA CN regulation.

The maintenance of mitochondria, including mtDNA replication and translation, is managed by a complex communication system between mtDNA and nDNA, which involves more than 1500 nDNA-encoded proteins (O'Brien, 2002). The system is still not fully understood, however several studies have identified important mtDNA replication machinery components such as *POLG*, *TFAM*, and *mtSSB* to be important for mtDNA CN regulation (Takamatsu *et al.*, 2002; Ruhanen *et al.*, 2010; Stewart *et al.*, 2011; Chinnery and Hudson, 2013). With so much still unknown about mtDNA CN regulation, I aimed to identify whether genetic factors modulate mtDNA CN changes in DNA sourced from white blood cell pellets in two large, healthy, and unrelated UK cohorts. In the process I have performed the largest, and most comprehensive, study of genetic, biological, and technical factors modulating mtDNA CN to date.

8.1 Genetic Factors Modulating Mitochondrial DNA Copy Number

8.1.1 Nuclear DNA and Epigenetic Factors

Two previous studies have attempted to identify genetic regulators of mtDNA CN, one performing linkage-analysis in Mexican-American families (Curran *et al.*, 2007), and a second using low coverage sequencing data to estimate mtDNA CN in a cohort of over 10,000 Chinese individuals (Cai *et al.*, 2015). Both studies used GWAS to identify loci associated with mtDNA CN changes. The linkage-analysis study identified several loci to be significantly associated with mtDNA CN regulation. These loci were involved in cellular trafficking, signalling, and metabolic events (Curran *et al.*, 2007). The GWAS in Chinese individuals did not replicate these results, and reported mitochondrial transcription factor A (*TFAM*), and cyclin-

dependent kinase 6 (*CDK6*) to be significantly associated with mtDNA CN changes (*p*= 8.73E-28, and *p*= 6.03E-16 respectively), suggesting mtDNA CN may be modulated by different loci. Furthermore, both studies had limitations; either using a limited number of polymorphic markers to map the genome, or relying on indirect measurements of mtDNA CN.

In an attempt to mitigate these limitations, but still identify genetic factors modulating mtDNA CN, I have performed a GWAS in over 8,000 people from two, independent, unrelated, healthy UK cohorts, using nearly 7.5 million imputed nuclear genetic markers to map loci associated to actual measures of mtDNA CN (0).

My data indicates that there are likely to be several loci regulating mtDNA CN. Analysis in each cohort revealed genome-wide significant (GWS) variants that were located in several different loci. Between cohort analyses, no variants or loci were replicated, and following meta-analysis of both cohorts, no variants were GWS.

We identified no GWS results (*p* <5.00E-08) as a consequence of considerable differences in phenotypic and experimental measurements between, and within UKBS and ALSPAC cohorts (Table 4.6 – section 4.4.4). Detailed analysis reported mtDNA CN proportions to change relative to DNA source (whole blood or white blood cell extracted), qPCR assay method, and blood cell component proportions. Restricting individuals, and adjusting for these factors in GWAS analysis weakened our study power, and as the data suggest, nDNA control of mtDNA CN is also likely to be complex. Our original study was powered to detect a low number of high effect alleles, however post-hoc power calculations, now based on actual data, determine that tens to hundreds of thousands of individuals will be needed to detect GWS loci associated with mtDNA CN changes with sufficient power, whilst accommodating to study heterogeneity.

Future work will need to ensure larger sample sizes are used, which have been extracted from the same DNA source, and have had their mtDNA CN quantified using the same qPCR assay, which is conducted by a single user. By applying these measures, relative mtDNA CN deviations will be minimised and will improve overall study power. Furthermore, a much more accurate and reproducible qPCR assay needs to be designed and optimised to minimise mtDNA CN variability between experimental runs, and across the general literature (explored in section 2.9). Additionally, if the DNA source is from blood, full blood count information will need to

be included and adjusted for in analysis to correct for cellular proportions. Following these procedures will minimise heterogeneity between cohorts, which we identified to affect GWAS results (Chapter 4).

Despite no results being definitive from this work, this is the most comprehensive study to investigate the effects cellular heterogeneity have on mtDNA CN, and GWAS results, and in relation to mtDNA CN regulation. We report gender-specific genetic effects on mtDNA CN regulation despite recording no significant difference in mtDNA CN between genders. This may be a result of the male nDNA, and likely the non-recombining Y chromosome, having to accommodate to newly inherited, also non-recombining, mtDNA following fertilisation (Al Rawi *et al.*, 2011; Sato and Sato, 2011; Politi *et al.*, 2014).

As previously discussed in section 4.1.1, the intergenic communication between nDNA and mtDNA is highly sensitive, and diversions from original nDNA-mtDNA communication makes it harder for tissues to generate sufficient ATP through OXPHOS (McKenzie *et al.*, 2004), which can alter tissue physical and metabolic performance because nDNA cannot recognise the inherited mtDNA (Nagao *et al.*, 1998; St. John, 2016). This may explain why mitochondrial biogenesis differences have been identified between genders in humans and rodents (Khalifa et al., 2017; Mauvais-Jarvis et al., 2017).

The lack of mtDNA CN difference reported between genders in this study is contradictory to the literature, because, although not investigated in this study, methylation of nDNA genes such as *POLGA* is likely to modulate mtDNA CN (Amaral *et al.*, 2007; Chinnery and Hudson, 2013; Lee *et al.*, 2015). However, Lopez and colleagues also reported different GWAS results between genders despite recording no significant differences in mtDNA CN, thus supporting our results. Suggesting sex-specific effects may be possible.

Elaborating further, mtDNA CN is directly controlled by methylation profiles of nDNA genes such as *POLGA* (Kelly *et al.*, 2012; Lee *et al.*, 2015), and methylation profiles are reported to change in blood leukocytes of obese individuals who are metabolically stressed (Wang *et al.*, 2010c). Furthermore, despite having low participant numbers in their study, Lynch and colleagues reported significantly lower natural killer cells, and cytotoxic T lymphocytes in obese individuals compared to lean individuals, however metabolically unhealthy obese individuals had significantly

lower natural killer cells than metabolically healthy obese individuals (Lynch *et al.*, 2009). Ineffective mitochondria can also directly contribute to blood composition changes (Pyle *et al.*, 2007; Pyle *et al.*, 2010; Jaffe and Irizarry, 2014), and cause anaemias because they prevent efficient erthyrocyte synthesis by interrupting iron metabolism, which depletes bone marrow stores and increases oxidative damage through the Fenton reaction (Ahlqvist *et al.*, 2015). This shows that blood cell composition changes dependent on mitochondrial metabolic stress, which could be managed by epigenetic programming.

Additionally, our GWAS performed on the healthy, female-only ALSPAC cohort reported multiple variants spanning *PSMD3* loci on chromosome 17 when not adjusting for blood cell composition. *PSMD3* is a gene which codes for a protein that is associated with neutrophil count. When blood cell composition was adjusted for, this locus association was removed, thus suggesting that mtDNA CN is relative to blood cell proportions. Similar findings have been reported in other studies (Pyle *et al.*, 2007; Pyle *et al.*, 2010; López *et al.*, 2012), which suggests that mtDNA CN may be a proxy to blood cell count, and that mtDNA CN may be regulated through complex feedback systems between metabolic stress signalling, which may trigger immune system intermediates, by acting through blood cell intermediates. This may explain why there were different genetic association results reported between gender-specific GWAS analysis. Epigenetic regulation may be contributing to differences reported between genders, because males are more susceptible to disease such as Parkinson ’s disease (PD) and amyotrophic lateral sclerosis (ALS) than females (Van Den Eeden et al., 2003; Kim et al., 2012), and females are more susceptible to autoimmune diseases than males (Whitacre, 2001; Fairweather and Rose, 2004), and these diseases are strongly determined by mitochondrial bioenergetics (Dupuis *et al.*, 2004; Pyle *et al.*, 2015a; Coxhead *et al.*, 2016).

Genetic and epigenetic factors have also been involved in autoimmune diseases (Farh *et al.*, 2015), and long-range interactions have been recorded between nDNA genes in autoimmune risk studies (Martin *et al.*, 2015). If there are complex systems interplaying between mitochondrial regulation, and immune intermediates which are being regulated by blood cell composition, then this may explain why several loci were reported in our nDNA GWAS, and it would be interesting to investigate whether they interacted on a chromosomal level, or through functional analysis in further studies.

It is important to consider whether mtDNA CN in blood is just an indicator of physiological cell stress, or is in fact just a mediator in triggering an immune response, or accommodating to the physiological change. For example, Kilbaugh and colleagues demonstrated that mtDNA CN significantly increased in peripheral blood from pigs when they had induced traumatic brain injury (Kilbaugh *et al.*, 2015). Also, when cells are stressed, cytokines are excreted, and mitochondria migrate between stressed cells to rescue aerobic respiration (Wang and Gerdes, 2015). Furthermore recent literature demonstrated that mitochondrial and ER interactions directly mediate the innate immune responses (Misawa *et al.*, 2017), and mitochondrial-ER interactions also directly mediate mtDNA CN (Friedman *et al.*, 2011; Lewis *et al.*, 2016).

Expanding on this further, we reported vesicle and peroxisome-related genes as top loci regions in our GWAS analysis to be associated with mtDNA CN, and peroxisomes, vesicle trafficking and ER maintenance are a tight-net signalling network for efficient lipid metabolism and cellular apoptosis regulation, which the mitochondria also oversee (Dimitrov *et al.*, 2013) (section 4.4.2). This suggests there may be an indirect association of functional genes contributing to mtDNA CN.

Furthermore, NUMTs have developed through evolution as a result of cellular stress, including ROS, which ensures mtDNA codes are selected for rather than relying on genetic drift (Ju *et al.*, 2015; Srinivasainagendra *et al.*, 2017). Literature is currently realising the extent the immune system cross talks between mtDNA and nDNA, and other GWAS have identified genes like interleukin-6 (*IL-6*), *IL-10*, superoxide dismutase-2 (*SOD2*), and insulin-like growth factor-1 (*IGF1*) to be associated with longevity (Garagnani *et al.*, 2014). mtDNA CN, and blood cell composition are reported to change with age (Niranjan *et al.*, 1982; Trifunovic *et al.*, 2004; Jaffe and Irizarry, 2014), which also suggests additional factors contributing to the complex interplay between mtDNA CN maintenance. This raises the possibility that immune factors could be contributing to mitochondrial stress over time, and the changes reported could be a result of natural ageing effects from ROS whereby ROS could be interrupting signalling events, to reduce the quality of mitochondria which would prevent them from functioning properly. Also, it is possible that ROS may be regulating cellular epigenetics which could result in increased immune stimulation, or increased autophagy. With ccf-mtDNA being increasingly recognised as a biomarker for disease, it raises the hypothesis that ccf-mtDNA could be an indicator of cell or

tissue stress with age which has led to disease, because ccf-mtDNA is generated from apoptotic, but not necrotic cells (Suzuki *et al.*, 2008).

Further work needs to be conducted on individual blood cell components to understand their genetic regulation with respect to mtDNA CN, because our work, and previous literature, reports that cells operate independently, and mtDNA CN is relative to their proportions or activity. We have identified the impact heterogeneity has on genetic association studies therefore, if work is going to be conducted in DNA from full blood, or white blood cell pellets, then full blood counts needs to be accommodated for to adjust for the cellular heterogeneity. As my work shows, not adjusting for cellular heterogeneity introduces results bias to more frequent cellular populations.

8.1.2 Mitochondrial DNA

Studies have reported differences in mitochondrial biogenesis and OXPHOS capacities between European mitochondrial haplogroups (Gomez-Duran *et al.*, 2010; Gómez-Durán *et al.*, 2012), and mtDNA CN is directly correlated to mitochondrial OXPHOS capacity (Dickinson *et al.*, 2013). Yet, this is the first, and largest, study to investigate whether mitochondrial DNA variants contribute to changes in mtDNA CN (Chapter 1). I performed GWAS across three, independent, unrelated cohorts using mitochondrial variants as a continuous quantitative trait. My data indicate that significantly lower mtDNA CN is present in individuals who have a mitochondrial haplogroup W background, when compared to the other European haplogroups. This result replicated in all three cohorts.

Haplogroup W is a relatively rare European haplogroup (around 3% prevalence). It is most prevalent in North-eastern countries such as Finland and Hungary, constituting 9.5% and 5% of their populations respectively. Due to its rarity, haplogroup W is often excluded as a group in studies making it difficult to compare to previous studies.

The two variants used to identify haplogroup W individuals in this study were *G5460A* and *G5046A*, which are located in the *MT-ND2* region either side of the O_L strand, and cause non-synonymous amino acid changes. Haplogroup WIX appears to be more common than haplogroup J in an aged Finnish population (Niemi et al., 2003), which either suggests haplogroup W predisposes individuals to age better than other haplogroups, or the study is biased because of population stratification numbers.

It could be possible that the non-synonymous amino acid changes in haplogroup W mtDNA may be altering the OXPHOS capacity of haplogroup W individuals. OXPHOS gene expression is known to change dependent on metabolic stress conditions from metabolites such as ROS (Morel and Barouki, 1999). With lower mtDNA CN, it is also possible to hypothesise that the mitochondria may generate less ROS if they are operating at a lower OXPHOS baseline threshold with fewer mtDNA molecules. If less ROS is being generated, or accumulated, from reduced OXPHOS then this may be preventing ageing events, which is resulting in higher numbers of individuals with haplogroup W mitochondria in the Finnish aged population.

Also, cytokine expression is dependent on mtDNA haplogroup variants (Bellizzi *et al.*, 2006), which suggests that immune response mediators are different between mitochondrial haplogroups. In this study, no significant difference in blood composition was reported between individuals who were haplogroup W or non-haplogroup W, which implies that the lower mtDNA CN reported are not an artefact of cellular proportions, or immune mediators. However, blood composition is reported to change depending on age and methylation profile (Houseman *et al.*, 2012), and CpG sites regulate methylation and cell composition. In mtDNA, CpG sites are located near O_L site, which is where the haplogroup W-defining variants are located. This evidence proposes that epigenetic mediators are controlling mtDNA CN relative to mitochondrial haplogroup variants.

Although not studied in this project, it is possible that epigenetics may be contributing to mtDNA CN regulation further. CpG islands are recognised to regulate the methylation patterns, which are critical for human induced pluripotent stem cells, and embryonic stem cells during differentiation. Cancers also have CpG distinguished shores (Doi *et al.*, 2009), and mtDNA CN changes are reported in cancers (Jiang *et al.*, 2006; Higuchi, 2007; Fan *et al.*, 2008; Malik and Czajka, 2013; Lee *et al.*, 2015). Mitochondrial haplogroups in embryonic stem cell models with the same nDNA background also have slightly different differential gene expression patterns, including *POLGA* (Kelly *et al.*, 2012; Kelly *et al.*, 2013a). This suggests mitochondrial haplogroups are affecting nDNA methylation profiles which may be managing mitochondrial replication and translation rates (St. John, 2016).

As the haplogroup W–defining variants in this study are located either side of the O_L site, it may be possible that epigenetic control over these variants may be interfering with the binding of replication machinery to the lagging strand, or are causing displacement of machinery in that region, which is contributing to decreased mtDNA synthesis (details in section 5.5.1).

Further work needs to be conducted to understand haplogroup W mitochondrial maintenance, because evidence indicates their physiology may explain a lot about overall mtDNA CN maintenance, and mitochondrial biogenesis. It would be beneficial to investigate the mechanisms by which these mitochondria function with lower mtDNA CN, as well as explore their tolerance to cellular stress, to further understand disease development in relation to mtDNA CN changes.

8.2 Biological Factors Modulating Mitochondrial DNA Copy Number

8.2.1 Age and Gender

Previous research suggests that ROS and oxidants significantly contribute to human ageing (Finkel and Holbrook, 2000; Liochev, 2013; Davalli *et al.*, 2016), and it is generally agreed that mtDNA CN decreases with age across a number of tissues as a result of ROS-induced damage (Tezze *et al.*; Barrientos *et al.*, 1997; Barazzoni *et al.*, 2000; Miller *et al.*, 2003; Cree *et al.*, 2008a; Cheng and Ivessa, 2010; Payne *et al.*, 2011; Mengel-From *et al.*, 2014; Laubenthal *et al.*, 2016; Zole *et al.*, 2017), however there is no conclusive evidence of any gender effects. This study reports statistically significant decreases in mtDNA CN with age, but only in males (section 3.3.4). Other studies have reported similar findings (Curran *et al.*, 2007), however these results may not be biologically signficiant because mtDNA CN decreases by 0.18 copies/cell every 10 years. Further work is needed to investigate the biological impact of this decrease in males with age.

Despite the potential lack of biological significance, these findings are important because they suggest that female mtDNA are less susceptible to oxidative damage than male mtDNA with age. Studies in mice have reported differences in mitochondrial biogenesis, respiration rate, ROS homeostasis, and morphology between males and females, and oxidative stress biomarkers are lower in females compared to age-matched males (Khalifa *et al.*, 2017). Additionally, females express higher levels of antioxidant genes than males (Borrás *et al.*, 2003). The results in this study agree with evidence in the literature, and also agree with Malik and colleagues

who hypothesised that mtDNA CN may change in response to oxidative stress (Malik *et al.*, 2011; Malik and Czajka, 2013).

Furthermore, mitochondrial-related diseases such as PD, ALS, and Leber's Hereditary Optic Neuropathy (LHON), which have reports of mtDNA CN changes (Wiedemann *et al.*, 2002; Giordano *et al.*, 2014; Pyle *et al.*, 2015a), are more predominant in males (Man *et al.*, 2003; Van Den Eeden *et al.*, 2003; Kim *et al.*, 2012). The penetrance of LHON has also been reduced as a outcome of increased mtDNA CN (Giordano et al., 2014), thereby suggesting there may be age-related gender-specific mechanisms impacting mtDNA CN levels. These compensatory mechanisms are important to understand because they may protect against mitochondrial genetic insults, and potentially provide a therapeutic target to prevent gender-specific metabolic age-related diseases.

In addition to this, a study demonstrated that oestrogen was an essential neuroprotective factor for age-related diseases such as ALS and LHON (Choi *et al.*, 2008; Kim *et al.*, 2012; Pisano *et al.*, 2015). This suggested that hormones may be protecting mtDNA from long-term oxidative damage.

8.2.2 Oestrogen-Independent Mitochondrial DNA Copy Number Regulation

Previous studies recognise oestrogen to have protective mechanisms on mitochondrial biogenesis through peroxisome proliferator-activated receptor (PPAR) pathways (Scarpulla, 2011; Kemper *et al.*, 2014), and is thought to contribute to the longer life expectancy observed in females (Viña *et al.*, 2005).

To expand on the age, and gender, specific mtDNA CN changes reported in section 3.3.4, and to investigate whether the changes were a result of oestrogen, I performed a study to investigate the differences in mtDNA CN between pre- and post-menopausal women (Chapter 6). The literature indicates that pre- and post-menopause can be accurately segregated using age (te Velde *et al.*, 1998; Wallace and Kelsey, 2010). However, stratification into pre- and post-menopausal women in this study (Chapter 6) identified no significant difference in mtDNA CN between groups, as well as no significant nDNA associations.

So how do we rationalise our lack of association? Pre- and post- menopausal women in this study were stratified by age, however stage and onset of menopause, as well as medications individuals were administering were not known. This restricted the

reliability of our results because population stratification was not accurate, and hormone therapies are reported to directly increase mtDNA CN (Knez *et al.*, 2016), which would have skewed our results further.

Needless to say, age is considered a good indicator of menopause, as the average female undergoes menopause by 50 years of age (te Velde *et al.*, 1998; Wallace and Kelsey, 2010). GWAS analysis in post-menopausal women reported several, consistent associations to variants which were located in *ABCG1, RAB27B* and *STARD6*, which are genes involved in phospholipid transport through vesicle-mediated pathways.

Overexpression of StARs is reported to decrease intracellular lipids, and lower inflammatory intermediates (Ning *et al.*, 2009). These results further highlight and support the results reported in the earlier nDNA meta-analysis work, whereby vesicle and lipid-related signalling events were associated with mtDNA CN changes through ER maintenance (Dimitrov *et al.*, 2013). If oestrogen levels decrease during menopause, it could be possible that the accumulation of lipids and sterols activate immune signalling pathways from increased mitochondrial stress, and ROS production. This increased physiological stress may be contributing to the development of menopause-related conditions such as osteoporosis, weight gain, hot flushes, disturbed sleep and skin atrophy.

The presence of oestrogen also increases mitochondrial respiration (Gigli and Bussmann, 2001; Nilsen *et al.*, 2006) which may suggest that oestrogen mediated pathways are contributing to downstream beta-oxidation, and lipid signalling and trafficking events, which are causing an increase in mitochondrial respiration. Further to the points highlighted previously, this proposes the hypothesis that mtDNA CN is just an indicator of respiration state, and the genetic maintenance of mtDNA CN is what dictates the capacity of mitochondria to accommodate to signalling pressures. The ability for mitochondria to accommodate to signalling pressures will then contribute to cellular stress signalling outcomes, which may activate the immune system. This may then explain why women are more likely to develop autoimmune diseases - as an indirect impact of oestrogen signalling overloading mitochondria, causing self-autophagy (Fairweather and Rose, 2004).

Further studies are needed to investigate oestrogen signalling and mtDNA CN regulation in more detail. However, exact stage of menopause, and medications

would need to be accounted for to ensure more precise analysis of mtDNA CN changes in pre- and post-menopausal women.

8.2.3 The Hereditability of Mitochondrial DNA Copy Number

Mitochondrial DNA is maternally inherited (Birky, 1978; Giles *et al.*, 1980), because paternal mtDNA is degraded soon after fertilisation (Al Rawi *et al.*, 2011; Sato and Sato, 2011; Politi *et al.*, 2014). Female oocytes contain both wildtype and mutated mtDNA (heteroplasmy) (Wallace and Chalkia, 2013; Hyslop *et al.*, 2016), which can be transmitted to the offspring. Heteroplasmy levels inherited dictate phenotype presentation, however this is dependent on tissue type and mtDNA mutation load (Sciacco *et al.*, 1994; Wallace and Chalkia, 2013; Stewart and Chinnery, 2015).

mtDNA CN is thought to regulate heteroplasmy transmission (Jenuth *et al.*, 1996; Cree *et al.*, 2008b), and mtDNA CN is reported to be between 33-65% inheritable in two independent studies in Mexican American families (Curran *et al.*, 2007), and monozygotic, dizygotic and sibling pairs (Xing *et al.*, 2008).

To further validate the heritability of mtDNA CN, I investigated mtDNA CN relationships in over 800 mother-child pairs (section 3.3.6). The data recorded contradict previous studies, because no correlation in mtDNA CN was recorded. However, I did not use any genetic data to perform linkage analysis between pairs, therefore limiting my results to not identify the exact heritability of mtDNA CN.

It would be beneficial to expand on my work in European populations, and identify the exact heritability of mtDNA CN in populations, and then compare this to mtDNA CN inheritance in other studies to identify if any genetic factors were responsible for any inheritance deviations.

8.3 Technical Factors Modulating Mitochondrial DNA Copy Number

Previous work conducted by Malik and colleagues demonstrated that it was important to design experiments correctly to determine mtDNA CN, to avoid bias or amplification of pseudoregions. Their work also identified that it was important to amplify non-variable mtDNA regions which are not present in nDNA pseudogenes, and highlighted the importance of correct DNA dilution series to ensure accurate Mt/N calculations (Malik *et al.*, 2011).

The work conducted in this study has expanded on Malik's work, and has investigated technical limitations of measuring mtDNA CN, and the effect this has on

large, population-based studies. From optimising an established qPCR method (Grady *et al.*, 2014a; Pyle *et al.*, 2015b; Rygiel *et al.*, 2015) in advance to conducting this study, I identified that triplicate and quadruplicate qPCR measurements were the most appropriate number of replicates to use, because more or less replicates resulted in increased standard deviations, which decrease genetic power in GWAS. Furthermore, I have recorded that there is no difference in mtDNA CN calculated using the average of triplicate Mt/N measurements, or using the median.

Furthermore, I have demonstrated why there are inconsistencies in mtDNA CN values reported across the literature. qPCR assay design, DNA handling procedures, and assay user all contribute to mtDNA CN variation within control samples (Chapter 4). Through collaboration with University of Bristol, whereby each cohort used a different qPCR assay to determine mtDNA CN, we identified a maximum Pearson's correlation value of 0.66 in mtDNA CN values generated from two different, established qPCR assays – one using TaqMan™ probes, and the other using SYBR® Green. Both methods use different technologies to determine mtDNA CN, and both produce different mtDNA CN results. These differences could be dependent on a number of things including DNA quality, DNA volume, and DNA concentration, but also the person processing the samples. However, these findings further support the need for an accurate and reproducible qPCR assay to be optimised to quantify mtDNA CN.

Additionally, by examining variants MAFs across mtDNA, I was able to identify polymorphic variants that may be interrupting established qPCR primer and probe binding domains (Chapter 1). By redesigning this assay, and using an alternative qPCR assay which quantified a less polymorphic region in *MT-RNR2* instead of *MT-ND1*, I was able to demonstrate that mtDNA CN values were much more consistent, and the ratio to *MT-ND4* was much more precise. This further identified the importance of needing to re-design established qPCR assays to detect both mtDNA CN, and mitochondrial deletions, because the current method of measuring the proportions of *MT-ND1* to *MT-ND4* results in false detections of mitochondrial deletions (Chapter 1).

I used LR-PCR methods to confirm that control samples which were originally identified to contain high levels of deletions (>20%) from qPCR analysis, actually contained no deletions (Chapter 1). Then, by comparing *MT-ND1* analysis with *MT-*

RNR2 analysis, indicated this may have been a result of variable *MT-ND1* values relative to *MT-ND4*, which could have been caused by a highly polymorphic variant being present beneath *MT-ND1* qPCR primer and probe binding domains. There are inconsistencies across the literature as to which reference genes should be used to quantify mtDNA CN, however the work here highlights the importance that qPCR design can contribute to inconsistent, and false positive results.

To compare mtDNA CN between studies, especially in genetic studies, mtDNA CN *must* be measured using the same qPCR assay, and ideally by the same user to minimise assay variability. Power calculations, where mtDNA CN standard deviations are high, decrease statistical power to identify genetic variants modulating mtDNA CN change (discussed in section 4.5.6.1). Therefore it is essential that qPCR assays are improved to minimise variation, and to improve precision of results reported across the literature.

Other methods of determining mtDNA CN are now starting to surface to minimise technical variations. Droplet-digital PCR (ddPCR) is more frequently being used because it is more accurate method to determine mtDNA CN as it amplifies single pieces of DNA using fluorescent probes to improve sensitivity and precision (Henrich *et al.*, 2012; Belmonte *et al.*, 2016; Ye *et al.*, 2017). Other studies are turning to genetic sequencing data to calculate mtDNA CN by calculating the ratio of aligned mitochondrial reads to aligned nuclear reads. These methods assume equal ploidy of both nuclear and mitochondrial regions of the genome but are becoming more popular as they minimise mtDNA CN technical variation between studies (Guo *et al.*, 2013; D'Erchia *et al.*, 2014; Cai *et al.*, 2015; Ding *et al.*, 2015; Reznik *et al.*, 2016). The main limitation to using genomic data to calculate mtDNA CN though is that it is not a direct measure of mtDNA CN in cells. For example, Cai and colleagues use proxy measurements of mtDNA CN to argue that mtDNA CN is a factor for major depressive disorder (MDD) (Cai *et al.*, 2015), however Tang and colleagues report mtDNA CN is not associated with MDD after using actual measures of mtDNA CN (He *et al.*, 2014).

Future work needs to investigate how genetic-calculated mtDNA CN values compare to actual measures of mtDNA CN before it can be assumed that methods are equivalent. However, a reliable, and concise qPCR first needs to be determined to compare to.

8.4 Conclusions

This study is the largest study to investigate genetic, biological, and technical factors modulating mtDNA CN across healthy, unrelated, European populations.

In conclusion, I report that:

8.4.1 Biological Conclusions

- There is no difference in mtDNA CN between males and females.
- mtDNA CN decreases with age, however this is significant in males, and not females.
- mtDNA CN changes in individual blood cell components, and this is dependent on peripheral blood cell proportions.
- Blood cell components are significantly different between males and females.
- mtDNA CN is not inherited maternally.
- No single nuclear-locus is regulating mtDNA CN, and there are indications that several loci are gender-specific.
- GWAS results are affected by phenotypic and experimental differences between cohorts; including DNA source, DNA extraction method, phenotype quantification methods, and blood cell proportions (when DNA is sourced from blood). All variables need to be adjusted for or minimised between cohorts.
- Epigenetic factors are likely to be contributing to mtDNA CN changes, and these may be influenced by mitochondrial-regulated metabolic activity, which is determined by mtDNA CN turnover.
- Our GWAS results indicate several genes that code for proteins involved in lipid signalling and vesicle-trafficking are associated to mtDNA CN regulation.
- mtDNA CN may change relative to blood cell proportions in response to cellular stress feedback mechanisms during immune response events.
- Mitochondrial variants effect mtDNA CN changes, and mitochondrial haplogroup W has significantly lower mtDNA CN compared to other European mitochondrial haplogroups.
- There is no significant difference in mtDNA CN between pre- and post-menopausal women, which suggests oestrogen is not mediating mtDNA numbers.
- Mitochondrial deletions are not present in healthy individual’s blood.

8.4.2 Technical Conclusions

- There is no difference in mtDNA CN calculated using mean or median values from replicates.
- Triplicate or quadruplicate replicates are the optimum number to use in qPCR assays.
- qPCR conducted by different users, or different qPCR assays (TaqMan™ vs SYBR® Green) produce inconsistent mtDNA CN values.
- mtDNA polymorphisms are altering the efficiency of qPCR primers and probes, resulting in inaccurate reporting of mtDNA CN values.
- *MT-RNR2* appears to be a more accurate mitochondrial region to quantify instead of *MT-ND1*.
- mtDNA CN is significantly different when DNA is extracted from whole blood, or white blood cell pellets, and analysis is not comparable between each DNA source.
- mtDNA CN methods need to be improved to find approaches that are comparable across the literature, but also representative of real biological measures of mtDNA CN in populations.
- There is very little difference in data imputed by the Sanger or Michigan Imputation Servers, and both datasets produce the same GWAS results.

8.5 Future Study Recommendations

As a result of this work, it is clear that much larger sample sizes are needed to investigate nuclear genetic factors modulating mtDNA CN in European populations. To minimise standard deviations, and prevent reduced statistical power, mtDNA CN must be extracted, stored, handled, and determined using the same procedures and quantitative assays.

In advance of further large mtDNA CN studies being conducted, it is imperative that established qPCR methods are revised, and optimised to measure mtDNA regions with few polymorphisms, which will cause technical inconsistencies. Additionally, several mtDNA regions need to be quantified in future qPCR assays to identify mtDNA quality, and to normalise regions relative to one another to ensure full, circular mtDNA is being quantified.

Furthermore, qPCR technologies report different mtDNA CN. To make data comparable across the literature, standardised methods need to be established and used.

mtDNA CN is different in individual blood cell components, which is dependent on overall cellular proportions. When investigating mtDNA CN in DNA extracted from blood samples, full blood count must be adjusted for. It would be interesting to separate blood cell components using fluorescence activated cell sorting technologies, and then perform GWAS in each cell component to identify genetic variants associated to mtDNA CN variation.

Also, performing RNA-sequencing in each blood cell component would be beneficial to identify genetic factors modulating the significantly lower mtDNA CN reported in mitochondria with haplogroup W mtDNA. Future work should investigate the physiological differences in haplogroup W mitochondria compared to other European haplogroups, and explore their OXPHOS capacity relative to their metabolic demands.

Further studies should be conducted to identify the inheritance of mtDNA CN in European populations, because current available literature is limited, and it would be beneficial to understand how mtDNA CN is inherited to predict disease susceptibility.

Finally, it would be interesting to further investigate the role oestrogen has on mtDNA CN regulation, especially following menopause. Another study would need to measure mtDNA CN in a larger cohort of females whose menopause stage is exactly determined, and can be stratified by any hormone-replacement therapies.

I propose that the quality of individual mitochondria, combined with epigenetic regulation of nDNA contribute to the mitochondria’s ability to manage cellular insult or stress. This in turn may contribute to metabolic differences, which then define mtDNA CN turnover through the mediation of mtDNA replication efficiency, which is affected by mtDNA variation. However extensive studies need to investigate this further.

Chapter 9 Bibliography

Aanen, D.K., Spelbrink, J.N. and Beekman, M. (2014) 'What cost mitochondria? The maintenance of functional mitochondrial DNA within and across generations', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1646), p. 20130438.

Abhishek, A., Bavishi, A., Bavishi, A. and Choudhary, M. (2011) 'Bacterial genome chimaerism and the origin of mitochondria', *Canadian Journal of Microbiology*, 57(1), pp. 49-61.

Acin-Perez, R., Russwurm, M., Gunnewig, K., Gertz, M., Zoidl, G., Ramos, L., Buck, J., Levin, L.R., Rassow, J., Manfredi, G. and Steegborn, C. (2011) 'A phosphodiesterase 2A isoform localized to mitochondria regulates respiration', *J Biol Chem*, 286(35), pp. 30423-32.

Ahlqvist, K.J., Leoncini, S., Pecorelli, A., Wortmann, S.B., Ahola, S., Forsstrom, S., Guerranti, R., De Felice, C., Smeitink, J., Ciccoli, L., Hamalainen, R.H. and Suomalainen, A. (2015) 'MtDNA mutagenesis impairs elimination of mitochondria during erythroid maturation leading to enhanced erythrocyte destruction', *Nat Commun*, 6, p. 6494.

Ahola-Olli, A.V., Wurtz, P., Havulinna, A.S., Aalto, K., Pitkanen, N., Lehtimaki, T., Kahonen, M., Lyytikainen, L.P., Raitoharju, E., Seppala, I., Sarin, A.P., Ripatti, S., Palotie, A., Perola, M., Viikari, J.S., Jalkanen, S., Maksimow, M., Salomaa, V., Salmi, M., Kettunen, J. and Raitakari, O.T. (2017) 'Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors', *Am J Hum Genet*, 100(1), pp. 40-50.

Aken, B.L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., García Girón, C., Hourlier, T., Howe, K., Kähäri, A., Kokocinski, F., Martin, F.J., Murphy, D.N., Nag, R., Ruffier, M., Schuster, M., Tang, Y.A., Vogel, J.-H., White, S., Zadissa, A., Flicek, P. and Searle, S.M.J. (2016) 'The Ensembl gene annotation system', *Database*, 2016, pp. baw093-baw093.

Al Rawi, S., Louvet-Vallée, S., Djeddi, A., Sachse, M., Culetto, E., Hajjar, C., Boyd, L., Legouis, R. and Galy, V. (2011) 'Postfertilization autophagy of sperm organelles prevents paternal mitochondrial DNA transmission', *Science*, 334(6059), pp. 1144-1147.

Albayrak, L., Khanipov, K., Pimenova, M., Golovko, G., Rojas, M., Pavlidis, I., Chumakov, S., Aguilar, G., Chávez, A., Widger, W.R. and Fofanov, Y. (2016) 'The ability of human nuclear DNA to cause false positive low-abundance heteroplasmy calls varies across the mitochondrial genome', *BMC Genomics*, 17(1), p. 1017.

Alberio, S., Mineri, R., Tiranti, V. and Zeviani, M. (2007) 'Depletion of mtDNA: Syndromes and genes', *Mitochondrion*, 7(1–2), pp. 6-12.

Albershardt, T.C., Iritani, B.M. and Ruddell, A. (2012) 'Evaluation of reference genes for quantitative PCR analysis of mouse lymphocytes', *Journal of immunological methods*, 384(1-2), pp. 196-199.

Allen, J.F. (2003) 'The function of genomes in bioenergetic organelles', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1429), pp. 19-38.

Amaral, A., Ramalho-Santos, J. and St John, J.C. (2007) 'The expression of polymerase gamma and mitochondrial transcription factor A and the regulation of mitochondrial DNA content in mature human sperm', *Human Reproduction*, 22(6), pp. 1585-1596.

Ameer, S.S., Xu, Y., Engström, K., Li, H., Tallving, P., Nermell, B., Boemo, A., Parada, L.A., Peñaloza, L.G., Concha, G., Harari, F., Vahter, M. and Broberg, K. (2016) 'Exposure to Inorganic Arsenic Is Associated with Increased Mitochondrial DNA Copy Number and Longer Telomere Length in Peripheral Blood', *Frontiers in Cell and Developmental Biology*, 4(87).

Ameur, A., Stewart, J.B., Freyer, C., Hagström, E., Ingman, M., Larsson, N.-G. and Gyllensten, U. (2011) 'Ultra-Deep Sequencing of Mouse Mitochondrial DNA: Mutational Patterns and Their Origins', *PLOS Genetics*, 7(3), p. e1002028.

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P. and Zondervan, K.T. (2010) 'Data quality control in genetic case-control association studies', *Nature Protocols*, 5(9), pp. 1564-1573.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R. and Young, I.G. (1981) 'Sequence and organization of the human mitochondrial genome', *Nature*, 290(5806), pp. 457-465.

Andersson, S.G.E., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C.M., Podowski, R.M., Näslund, A.K., Eriksson, A.S., Winkler, H.H. and Kurland, C.G. (1998) 'The genome sequence of Rickettsia prowazekii and the origin of mitochondria', *Nature*, 396(6707), pp. 133-140.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. and Howell, N. (1999) 'Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA', *Nature Genetics*, 23(2), p. 147.

Applied Biosystems (2002) 'SYBR Green PCR Master Mix and RT-PCR Protocol', [Online].

Arnal, J.-F., Lenfant, F., Metivier, R., Flouriot, G., Henrion, D., Adlanmerini, M., Fontaine, C., Gourdy, P., Chambon, P., Katzenellenbogen, B. and Katzenellenbogen, J. (2017) *Membrane and Nuclear Estrogen Receptor Alpha Actions: From Tissue Specificity to Medical Implications*.

Audet-walsh, E. and Giguere, V. (2015) 'The multiple universes of estrogen-related receptor [alpha] and [gamma] in metabolic control and related diseases', *Acta Pharmacol Sin*, 36(1), pp. 51-61.

Aulchenko, Y.S., Struchalin, M.V., Belonogova, N.M., Axenovich, T.I., Weedon, M.N., Hofman, A., Uitterlinden, A.G., Kayser, M., Oostra, B.A., van Duijn, C.M., Janssens, A.C.J.W. and Borodin, P.M. (2009) 'Predicting human height by Victorian and genomic methods', *Eur J Hum Genet*, 17(8), pp. 1070-1075.

Bain, B., Seed, M. and Godsland, I. (1984) 'Normal values for peripheral blood white cell counts in women of four different ethnic origins', *Journal of Clinical Pathology*, 37(2), pp. 188-193.

Bain, B.J. (1996) 'Ethnic and sex differences in the total and differential white cell count and platelet count', *Journal of Clinical Pathology*, 49(8), pp. 664-666.

Baker, M.J., Frazier, A.E., Gulbis, J.M. and Ryan, M.T. (2007) 'Mitochondrial protein-import machinery: correlating structure with function', *Trends in Cell Biology*, 17(9), pp. 456-464.

Balch, C.M., Buzaid, A.C., Soong, S.-J., Atkins, M.B., Cascinelli, N., Coit, D.G., Fleming, I.D., Gershenwald, J.E., Houghton, A., Kirkwood, J.M., McMasters, K.M.,

Mihm, M.F., Morton, D.L., Reintgen, D.S., Ross, M.I., Sober, A., Thompson, J.A. and Thompson, J.F. (2001) 'Final Version of the American Joint Committee on Cancer Staging System for Cutaneous Melanoma', *Journal of Clinical Oncology*, 19(16), pp. 3635-3648.

Ballard, J.W.O., Katewa, S.D., Melvin, R.G., Chan, G. and Ballard, B. 1114 (2007) 'Comparative analysis of mitochondrial genotype and aging' *Annals of the New York Academy of Sciences*. pp. 93-106.

Barazzoni, R., Short, K.R. and Nair, K.S. (2000) 'Effects of Aging on Mitochondrial DNA Copy Number and Cytochromec Oxidase Gene Expression in Rat Skeletal Muscle, Liver, and Heart', *Journal of Biological Chemistry*, 275(5), pp. 3343-3347.

Barrientos, A., Casademont, J., Cardellach, F., Estivill, X., Urbano-Marquez, A. and Nunes, V. (1997) 'Reduced steady-state levels of mitochondrial RNA and increased mitochondrial DNA amount in human brain with aging', *Brain Res Mol Brain Res*, 52.

Barsh, G.S., Copenhaver, G.P., Gibson, G. and Williams, S.M. (2012) 'Guidelines for Genome-Wide Association Studies', *PLOS Genetics*, 8(7), p. e1002812.

Bartlett, K. and Eaton, S. (2004) 'Mitochondrial β-oxidation', *European Journal of Biochemistry*, 271(3), pp. 462-469.

Beckman Coulter Inc. (2011a) 'Operation Principles', in Inc, B.C. (ed.) *COULTER HmX Hematology Analyzer - Reference*. Beckman Coulter Inc., pp. 35-52.

Beckman Coulter Inc. (2011b) 'Operation Principles: Parameters and Their Derivation', in Inc, B.C. (ed.) *COULTER HmX Hematology Analyzer - Reference*. Beckman Coulter Inc., pp. 48-51.

Beinert, H., Holm, R.H. and Münck, E. (1997) 'Iron-Sulfur Clusters: Nature's Modular, Multipurpose Structures', *Science*, 277(5326), pp. 653-659.

Beisswenger, P.J., Drummond, K.S., Nelson, R.G., Howell, S.K., Szwergold, B.S. and Mauer, M. (2005) 'Susceptibility to Diabetic Nephropathy Is Related to Dicarbonyl and Oxidative Stress', *Diabetes*, 54(11), pp. 3274-3281.

Bellamy, G.J., Hinchliffe, R.F., Crawshaw, K.C., Finn, A. and Bell, F. (2000) 'Total and differential leucocyte counts in infants at 2, 5 and 13 months of age', *Clinical & Laboratory Haematology*, 22(2), pp. 81-87.

Bellizzi, D., Cavalcante, P., Taverna, D., Rose, G., Passarino, G., Salvioli, S., Franceschi, C. and De Benedictis, G. (2006) 'Gene expression of cytokines and cytokine receptors is modulated by the common variability of the mitochondrial DNA in cybrid cell lines', *Genes to Cells*, 11(8), pp. 883-891.

Belmont, J.W., Boudreau, A., Leal, S.M., Hardenbol, P., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhou, J., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Parkin, M., Roy, J., Stahl, E., Winchester, E., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M.Y., Tsui, S.K.W., Xue, H., Wong, J.T.F., Galver, L.M., Fan, J.B., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallée, C., Verner, A., Hudson, T.J., Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Sham, P.C., Song, Y.Q., Tam, P.K.H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., et al. (2005) 'A haplotype map of the human genome', *Nature*, 437(7063), pp. 1299-1320.

Belmonte, F.R., Martin, J.L., Frescura, K., Damas, J., Pereira, F., Tarnopolsky, M.A. and Kaufman, B.A. (2016) 'Digital PCR methods improve detection sensitivity and measurement precision of low abundance mtDNA deletions', *Scientific Reports*, 6, p. 25186.

Bender, A., Krishnan, K.J., Morris, C.M., Taylor, G.A., Reeve, A.K., Perry, R.H., Jaros, E., Hersheson, J.S., Betts, J., Klopstock, T., Taylor, R.W. and Turnbull, D.M. (2006) 'High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease', *Nature Genetics*, 38(5), pp. 515-517.

Berger, J. and Moller, D.E. 53 (2002) 'The mechanisms of action of PPARs' *Annual Review of Medicine*. pp. 409-435.

Billings, L.K. and Florez, J.C. (2010) 'The genetics of type 2 diabetes: what have we learned from GWAS?', *Annals of the New York Academy of Sciences*, 1212, pp. 59-77.

Birky, C.W. (1978) 'Transmission Genetics of Mitochondria and Chloroplasts', *Annual Review of Genetics*, 12(1), pp. 471-512.

Blake, J.C., Taanman, J.-W., Morris, A.M.M., Gray, R.G.F., Cooper, J.M., McKiernan, P.J., Leonard, J.V. and Schapira, A.H.V. (1999) 'Mitochondrial DNA Depletion Syndrome is Expressed in Amniotic Fluid Cell Cultures', *The American Journal of Pathology*, 155(1), pp. 67-70.

Blanchard, J.L. and Lynch, M. (2000) 'Organellar genes', *Trends in Genetics*, 16(7), pp. 315-320.

Bogenhagen, D. and Clayton, D.A. (1977) 'Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle', *Cell*, 11(4), pp. 719-727.

Bogenhagen, D.F., Rousseau, D. and Burke, S. (2008) 'The layered structure of human mitochondrial DNA nucleoids', *Journal of Biological Chemistry*, 283(6), pp. 3665-3675.

Boldogh, I.R. and Pon, L.A. (2006) 'Interactions of mitochondria with the actin cytoskeleton', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1763(5–6), pp. 450-462.

Bonod-Bidaud, C., Giraud, S., Mandon, G., Mousson, B. and Stepien, G. (1999) 'Quantification of OXPHOS Gene Transcripts during Muscle Cell Differentiation in Patients with Mitochondrial Myopathies', *Experimental Cell Research*, 246(1), pp. 91-97.

Borrás, C., Gambini, J., Gómez-Cabrera, M.C., Sastre, J., Pallardó, F.V., Mann, G.E. and Viña, J. (2006) 'Genistein, a soy isoflavone, up-regulates expression of antioxidant genes: involvement of estrogen receptors, ERK1/2, and NFκB', *The FASEB Journal*, 20(12), pp. 2136-2138.

Borrás, C., Gambini, J., López-Grueso, R., Pallardó, F.V. and Viña, J. (2010) 'Direct antioxidant and protective effect of estradiol on isolated mitochondria', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1802(1), pp. 205-211.

Borrás, C., Sastre, J., García-Sala, D., Lloret, A., Pallardó, F.V. and Viña, J. (2003) 'Mitochondria from females exhibit higher antioxidant gene expression and lower oxidative damage than males', *Free Radical Biology and Medicine*, 34(5), pp. 546-552.

Bourgeois, J.M. and Tarnopolsky, M.A. (2004) 'Pathology of skeletal muscle in mitochondrial disorders', *Mitochondrion*, 4(5–6), pp. 441-452.

Brand, K. (1997) 'Aerobic glycolysis by proliferating cells: Protection against oxidative stress at the expense of energy yield', *Journal of Bioenergetics and Biomembranes*, 29(4), pp. 355-364.

Brandon, M.C., Lott, M.T., Nguyen, K.C., Spolim, S., Navathe, S.B., Baldi, P. and Wallace, D.C. (2005) 'MITOMAP: a human mitochondrial genome database—2004 update', *Nucleic Acids Research*, 33(Database Issue), pp. D611-D613.

Breuer, M.E., Koopman, W.J., Koene, S., Nooteboom, M., Rodenburg, R.J., Willems, P.H. and Smeitink, J.A.M. (2013) 'The role of mitochondrial OXPHOS dysfunction in the development of neurologic diseases', *Neurobiology of Disease*, 51, pp. 27-34.

Broekmans, F.J.M., J Faddy, M., Scheffer, G.J. and te Velde, E. (2004) *Antral follicle counts are related to age at natural fertility loss and age at menopause*.

Brown, W.M., George Jr, M. and Wilson, A.C. (1979) 'Rapid evolution of animal mitochondrial DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 76(4), pp. 1967-1971.

Browning, S.R. and Browning, B.L. (2011) 'Haplotype phasing: existing methods and new developments', *Nat Rev Genet*, 12(10), pp. 703-714.

Brunstein J. (2015) 'Freeze-thaw cycles and nucleic acid stability: what's safe for your samples?', *Medical Laboratory Observer*, [Online]. Available at: https://www.mlo-online.com/freeze-thaw-cycles-and-nucleic-acid-stability-whats-safe-for-your-samples.php (Accessed: July 21, 2017).

Buler, M., Aatsinki, S.M., Izzi, V., Uusimaa, J. and Hakkola, J. (2014) 'SIRT5 is under the control of PGC-1o: And AMPK and is involved in regulation of mitochondrial energy metabolism', *FASEB Journal*, 28(7), pp. 3225-3237.

Burger, H.G. (1996) 'The endocrinology of the menopause', *Maturitas*, 23(2), pp. 129-136.

Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W.H., Samani, N.J., Todd, J.A., Donnelly, P., Barrett, J.C., Davison, D., Easton, D., Evans, D., Leung, H.T., Marchini, J.L., Morris, A.P., Spencer, C.C.A., Tobin, M.D., Attwood, A.P., Boorman, J.P., Cant, B., Everson, U., Hussey, J.M., Jolley, J.D., Knight, A.S., Koch, K., Meech, E., Nutland, S., Prowse, C.V., Stevens, H.E., Taylor, N.C., Walters, G.R., Walker, N.M., Watkins, N.A., Winzer, T., Jones, R.W., McArdle, W.L., Ring, S.M., Strachan, D.P., Pembrey, M., Breen, G., St. Clair, D., Caesar, S., Gordon-Smith, K., Jones, L., Fraser, C., Green, E.K., Grozeva, D., Hamshere, M.L., Holmans, P.A., Jones, I.R., Kirov, G., Moskvina, V., Nikolov, I., O'Donovan, M.C., Owen, M.J., Collier, D.A., Elkin, A., Farmer, A., Williamson, R., McGuffin, P., Young, A.H., Ferrier, I.N., Ball, S.G., Balmforth, A.J., Barrett, J.H., Bishop, D.T., Iles, M.M., Maqbool, A., Yuldasheva, N., Hall, A.S., Braund, P.S., Dixon, R.J., Mangino, M., Stevens, S., Thompson, J.R., Bredin, F., Tremelling, M., Parkes, M., Drummond, H., Lees, C.W., Nimmo, E.R., Satsangi, J., Fisher, S.A., Forbes, A., Lewis, C.M., Onnie, C.M., Prescott, N.J., Sanderson, J., Mathew, C.G., Barbour, J., Mohiuddin, M.K., Todhunter, C.E., Mansfield, J.C., Ahmad, T., Cummings, F.R., Jewell, D.P., et al. (2007) 'Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls', *Nature*, 447(7145), pp. 661-678.

Cai, N., Li, Y., Chang, S., Liang, J., Lin, C., Zhang, X., Liang, L., Hu, J., Chan, W., Kendler, K.S., Malinauskas, T., Huang, G.J., Li, Q., Mott, R. and Flint, J. (2015) 'Genetic Control over mtDNA and Its Relationship to Major Depressive Disorder', *Current Biology*, 25(24), pp. 3170-3177.

Caielli, S., Athale, S., Domic, B., Murat, E., Chandra, M., Banchereau, R., Baisch, J., Phelps, K., Clayton, S., Gong, M., Wright, T., Punaro, M., Palucka, K., Guiducci, C., Banchereau, J. and Pascual, V. (2016) 'Oxidized mitochondrial nucleoids released by neutrophils drive type I interferon production in human lupus', *The Journal of Experimental Medicine*, 213(5), pp. 697-713.

Calabrese, F.M., Simone, D. and Attimonelli, M. (2012) 'Primates and mouse NumtS in the UCSC Genome Browser', *BMC Bioinform*, 13 Suppl 4.

Calvo, S.E., Clauser, K.R. and Mootha, V.K. (2016) 'MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins', *Nucleic Acids Res*, 44(D1), pp. D1251-7.

Cámara, Y., González-Vioque, E., Scarpelli, M., Torres-Torronteras, J., Caballero, A., Hirano, M. and Martí, R. (2014) 'Administration of deoxyribonucleosides or inhibition of their catabolism as a pharmacological approach for mitochondrial DNA depletion syndrome', *Human Molecular Genetics*, 23(9), pp. 2459-2467.

Camus, M.F., Clancy, David J. and Dowling, Damian K. (2012) 'Mitochondria, Maternal Inheritance, and Male Aging', *Current Biology*, 22(18), pp. 1717-1721.

Cann, R.L., Stoneking, M. and Wilson, A.C. (1987) 'Mitochondrial DNA and human evolution', *Nature*, 325(6099), pp. 31-36.

Cao, L., Shitara, H., Horii, T., Nagao, Y., Imai, H., Abe, K., Hara, T., Hayashi, J.-I. and Yonekawa, H. (2007) 'The mitochondrial bottleneck occurs without reduction of mtDNA content in female mouse germ cells', *Nat Genet*, 39(3), pp. 386-390.

Cao, L., Shitara, H., Sugimoto, M., Hayashi, J.-I., Abe, K. and Yonekawa, H. (2009) 'New Evidence Confirms That the Mitochondrial Bottleneck Is Generated without Reduction of Mitochondrial DNA Content in Early Primordial Germ Cells of Mice', *PLoS Genetics*, 5(12), p. e1000756.

Capaldi, R.A., Murray, J., Byrne, L., Janes, M.S. and Marusich, M.F. (2004) 'Immunological approaches to the characterization and diagnosis of mitochondrial disease', *Mitochondrion*, 4(5–6), pp. 417-426.

Capps, G.J., Samuels, D.C. and Chinnery, P.F. (2003) 'A model of the nuclear control of mitochondrial DNA replication', *J Theor Biol*, 221(4), pp. 565-83.

Carling, P.J., Cree, L.M. and Chinnery, P.F. (2011) 'The implications of mitochondrial DNA copy number regulation during embryogenesis', *Mitochondrion*, 11(5), pp. 686-692.

Cerezo, M., Achilli, A., Olivieri, A., Perego, U.A., Gómez-Carballa, A., Brisighelli, F., Lancioni, H., Woodward, S.R., López-Soto, M., Carracedo, Á., Capelli, C., Torroni, A.

and Salas, A. (2012) 'Reconstructing ancient mitochondrial DNA links between Africa and Europe', *Genome Research*, 22(5), pp. 821-826.

Chagnon, P., Gee, M., Filion, M., Robitaille, Y., Belouchi, M. and Gauvreau, D. (1999) 'Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with Alzheimer disease and controls in a French-Canadian founder population', *American Journal of Medical Genetics*, 85(1), pp. 20-30.

Chan, D.C. 46 (2012) 'Fusion and fission: Interlinked processes critical for mitochondrial health' *Annual Review of Genetics* [Article]. pp. 265-287.

Chan, S.W., Chevalier, S., Aprikian, A. and Chen, J.Z. (2013) 'Simultaneous Quantification of Mitochondrial DNA Damage and Copy Number in Circulating Blood: A Sensitive Approach to Systemic Oxidative Stress', *BioMed Research International*, 2013, p. 10.

Chang, D.D. and Clayton, D.A. (1985) 'Priming of human mitochondrial DNA replication occurs at the light-strand promoter', *Proceedings of the National Academy of Sciences of the United States of America*, 82(2), pp. 351-355.

Chang, X., Wang, Z., Hao, P., Li, Y.Y. and Li, Y.X. (2010) 'Exploring mitochondrial evolution and metabolism organization principles by comparative analysis of metabolic networks', *Genomics*, 95(6), pp. 339-344.

Chaube, B. and Bhat, M.K. (2016) 'AMPK, a key regulator of metabolic/energy homeostasis and mitochondrial biogenesis in cancer cells', *Cell Death & Disease*, 7(1), p. e2044.

Chen, H., Vermulst, M., Wang, Y.E., Chomyn, A., Prolla, T.A., McCaffery, J.M. and Chan, D.C. (2010) 'Mitochondrial fusion is required for mtDNA stability in skeletal muscle and tolerance of mtDNA mutations', *Cell*, 141(2), pp. 280-289.

Chen, J.-Q., Cammarata, P.R., Baines, C.P. and Yager, J.D. (2009) 'Regulation of mitochondrial respiratory chain biogenesis by estrogens/estrogen receptors and physiological, pathological and pharmacological implications', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1793(10), pp. 1540-1570.

Chen, J.B., Lin, T.K., Liou, C.W., Liao, S.C., Lee, L.C., Wang, P.W. and Tiao, M.M. (2008) 'Correlation of Oxidative Stress Biomarkers and Peritoneal Urea Clearance with Mitochondrial DNA Copy Number in Continuous Ambulatory Peritoneal Dialysis Patients', *American Journal of Nephrology*, 28(5), pp. 853-859.

Chen, T., He, J., Shen, L., Fang, H., Nie, H. and Jin, T. (2011) 'The mitochondrial DNA 4,977-bp deletion and its implication in copy number alteration in colorectal cancer', *BMC Med Genet*, 12.

Chen, X., Prosser, R., Simonetti, S., Sadlock, J., Jagiello, G. and Schon, E.A. (1995) 'Rearranged mitochondrial genomes are present in human oocytes', *American Journal of Human Genetics*, 57(2), pp. 239-247.

Cheng, X. and Ivessa, A.S. (2010) 'The migration of mitochondrial DNA fragments to the nucleus affects the chronological aging process of Saccharomyces cerevisiae', *Aging Cell*, 9.

Cherry, A.D. and Piantadosi, C.A. (2015) 'Regulation of Mitochondrial Biogenesis and Its Intersection with Inflammatory Responses', *Antioxidants & Redox Signaling*, 22(12), pp. 965-976.

Chiang, L., Ngo, J., Schechter, J.E., Karvar, S., Tolmachova, T., Seabra, M.C., Hume, A.N. and Hamm-Alvarez, S.F. (2011) 'Rab27b regulates exocytosis of secretory vesicles in acinar epithelial cells from the lacrimal gland', *American Journal of Physiology - Cell Physiology*, 301(2), pp. C507-C521.

Chinnery, P.F., Craven, L., Mitalipov, S., Stewart, J.B., Herbert, M. and Turnbull, D.M. (2014) 'The Challenges of Mitochondrial Replacement', *PLoS Genetics*, 10(4), p. e1004315.

Chinnery, P.F., DiMauro, S., Shanske, S., Schon, E.A., Zeviani, M., Mariotti, C., Carrara, F., Lombes, A., Laforet, P., Ogier, H., Jaksch, M., Lochmüller, H., Horvath, R., Deschauer, M., Thorburn, D.R., Bindoff, L.A., Poulton, J., Taylor, R.W., Matthews, J.N.S. and Turnbull, D.M. (2004) 'Risk of developing a mitochondrial DNA deletion disorder', *The Lancet*, 364(9434), pp. 592-596.

Chinnery, P.F., Elliott, H.R., Hudson, G., Samuels, D.C. and Relton, C.L. (2012) 'Epigenetics, epidemiology and mitochondrial DNA diseases', *International Journal of Epidemiology*, 41(1), pp. 177-187.

Chinnery, P.F., Elliott, H.R., Syed, A., Rothwell, P.M. and for the Oxford Vascular, S. (2010) 'Mitochondrial DNA haplogroups and risk of transient ischaemic attack and ischaemic stroke: a genetic association study', *Lancet Neurology*, 9(5), pp. 498-503.

Chinnery, P.F., Howell, N., Lightowlers, R.N. and Turnbull, D.M. (1997) 'Molecular pathology of MELAS and MERRF. The relationship between mutation load and clinical phenotypes', *Brain*, 120(10), pp. 1713-1721.

Chinnery, P.F. and Hudson, G. (2013) 'Mitochondrial genetics', *British Medical Bulletin*, 106(1), pp. 135-159.

Chinnery, P.F., Taylor, D.J., Manners, D., Styles, P. and Lodi, R. (2001) 'No correlation between muscle A3243G mutation load and mitochondrial function in vivo', *Neurology*, 56(8), pp. 1101-1104.

Cho, S.B., Koh, I., Nam, H.-Y., Jeon, J.-P., Lee, H.K. and Han, B.-G. (2017) 'Mitochondrial DNA copy number augments performance of A1C and oral glucose tolerance testing in the prediction of type 2 diabetes', *Scientific Reports*, 7, p. 43203.

Choi, C.-I., Lee, Y.-D., Gwag, B.J., Cho, S.I., Kim, S.-S. and Suh-Kim, H. (2008) 'Effects of estrogen on lifespan and motor functions in female hSOD1 G93A transgenic mice', *Journal of the Neurological Sciences*, 268(1), pp. 40-47.

Clark, K.A., Howe, D.K., Gafner, K., Kusuma, D., Ping, S., Estes, S. and Denver, D.R. (2012) 'Selfish Little Circles: Transmission Bias and Evolution of Large Deletion-Bearing Mitochondrial DNA in Caenorhabditis briggsae Nematodes', *PLOS ONE*, 7(7), p. e41433.

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

Clayton, D.A. (1982) 'Replication of animal mitochondrial DNA', *Cell*, 28(4), pp. 693-705.

Clayton, D.A. (1987) 'Nuclear gene products that function in mitochondrial DNA replication', *Philosophical transactions of the Royal Society of London. Series B: Biological sciences*, 317(1187), pp. 473-482.

Coffee, Caffeine Genetics, C., Cornelis, M.C., Byrne, E.M., Esko, T., Nalls, M.A., Ganna, A., Paynter, N., Monda, K.L., Amin, N., Fischer, K., Renstrom, F., Ngwa, J.S., Huikari, V., Cavadino, A., Nolte, I.M., Teumer, A., Yu, K., Marques-Vidal, P., Rawal, R., Manichaikul, A., Wojczynski, M.K., Vink, J.M., Zhao, J.H., Burlutsky, G., Lahti, J., Mikkilä, V., Lemaitre, R.N., Eriksson, J., Musani, S.K., Tanaka, T., Geller, F., Luan, J.a., Hui, J., Mägi, R., Dimitriou, M., Garcia, M.E., Ho, W.-K., Wright, M.J., Rose, L.M., Magnusson, P.K.E., Pedersen, N.L., Couper, D., Oostra, B.A., Hofman, A., Ikram, M.A., Tiemeier, H.W., Uitterlinden, A.G., van Rooij, F.J.A., Barroso, I., Johansson, I., Xue, L., Kaakinen, M., Milani, L., Power, C., Snieder, H., Stolk, R.P., Baumeister, S.E., Biffar, R., Gu, F., Bastardot, F., Kutalik, Z., Jacobs, D.R., Forouhi, N.G., Mihailov, E., Lind, L., Lindgren, C., Michaëlsson, K., Morris, A., Jensen, M., Khaw, K.-T., Luben, R.N., Wang, J.J., Männistö, S., Perälä, M.-M., Kähönen, M., Lehtimäki, T., Viikari, J., Mozaffarian, D., Mukamal, K., Psaty, B.M., Döring, A., Heath, A.C., Montgomery, G.W., Dahmen, N., Carithers, T., Tucker, K.L., Ferrucci, L., Boyd, H.A., Melbye, M., Treur, J.L., Mellström, D., Hottenga, J.J., Prokopenko, I., Tönjes, A., Deloukas, P., Kanoni, S., Lorentzon, M., Houston, D.K., Liu, Y., et al. (2015) 'Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption', *Molecular psychiatry*, 20(5), pp. 647-656.

Cogswell, A.M., Stevens, R.J. and Hood, D.A. (1993) 'Properties of skeletal muscle mitochondria isolated from subsarcolemmal and intermyofibrillar regions', *American Journal of Physiology - Cell Physiology*, 264(2 33-2).

Collombet, J.M., Faure-Vigny, H., Mandon, G., Dumoulin, R., Boissier, S., Bernard, A., Mousson, B. and Stepien, G. (1997) 'Expression of oxidative phosphorylation genes in muscle cell cultures from patients with mitochondrial myopathies', *Mol Cell Biochem*, 168(1-2), pp. 73-85.

consortium, C. (2015) 'Sparse whole-genome sequencing identifies two loci for major depressive disorder', *Nature*, 523(7562), pp. 588-91.

Cooper, G.M. and Shendure, J. (2011) 'Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data', *Nat Rev Genet*, 12(9), pp. 628-640.

Copeland, W.C. (2012) 'Defects in mitochondrial DNA replication and human disease', *Crit Rev Biochem Mol Biol*, 47(1), pp. 64-74.

Corrêa, T.D., Pereira, A.J., Brandt, S., Vuda, M., Djafarzadeh, S., Takala, J. and Jakob, S.M. (2017) 'Time course of blood lactate levels, inflammation, and mitochondrial function in experimental sepsis', *Critical Care*, 21(1), p. 105.

Coskun, P.E., Ruiz-Pesini, E. and Wallace, D.C. (2003) 'Control region mtDNA variants: Longevity, climatic adaptation, and a forensic conundrum', *Proceedings of the National Academy of Sciences of the United States of America*, 100(5), pp. 2174-2176.

Coulam, C.B., Adamson, S.C. and Annegers, J.F. (1986) 'Incidence of premature ovarian failure', *Obstet Gynecol*, 67(4), pp. 604-6.

Coxhead, J., Kurzawa-Akanbi, M., Hussain, R., Pyle, A., Chinnery, P. and Hudson, G. (2016) 'Somatic mtDNA variation is an important component of Parkinson's disease', *Neurobiol Aging*, 38.

Cramer, D.W., Xu, H. and Harlow, B.L. (1995) 'Family history as a predictor of early menopause*', *Fertility and Sterility*, 64(4), pp. 740-745.

Craven, L., Tang, M.-X., Gorman, G.S., De Sutter, P. and Heindryckx, B. (2017) 'Novel reproductive technologies to prevent mitochondrial disease', *Human Reproduction Update*.

Cree, L.M., Patel, S.K., Pyle, A., Lynn, S., Turnbull, D.M., Chinnery, P.F. and Walker, M. (2008a) 'Age-related decline in mitochondrial DNA copy number in isolated human pancreatic islets', *Diabetologia*, 51(8), pp. 1440-1443.

Cree, L.M., Samuels, D.C., de Sousa Lopes, S.C., Rajasimha, H.K., Wonnapinij, P., Mann, J.R., Dahl, H.-H.M. and Chinnery, P.F. (2008b) 'A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes', *Nat Genet*, 40(2), pp. 249-254.

Crimmins, D., Morris, J.G.L., Walker, G.L., Sue, C.M., Byrne, E., Stevens, S., Jean-Francis, B., Yiannikas, C. and Pamphlett, R. (1993) 'Mitochondrial encephalomyopathy: Variable clinical expression within a single kindred', *Journal of Neurology Neurosurgery and Psychiatry*, 56(8), pp. 900-905.

Cristina Kenney, M., Chwa, M., Atilano, S.R., Falatoonzadeh, P., Ramirez, C., Malik, D., Tarek, M., Cáceres-del-Carpio, J., Nesburn, A.B., Boyer, D.S., Kuppermann,

B.D., Vawter, M., Michal Jazwinski, S., Miceli, M., Wallace, D.C. and Udar, N. (2014) 'Inherited mitochondrial DNA variants can affect complement, inflammation and apoptosis pathways: Insights into mitochondrial-nuclear interactions', *Human Molecular Genetics*, 23(13), pp. 3537-3551.

Crompton, M., Virji, S. and Ward, J.M. (1998) 'Cyclophilin-D binds strongly to complexes of the voltage-dependent anion channel and the adenine nucleotide translocase to form the permeability transition pore', *European Journal of Biochemistry*, 258(2), pp. 729-735.

Curran, J.E., Johnson, M.P., Dyer, T.D., Göring, H.H.H., Kent, J.W., Charlesworth, J.C., Borg, A.J., Jowett, J.B.M., Cole, S.A., MacCluer, J.W., Kissebah, A.H., Moses, E.K. and Blangero, J. (2007) 'Genetic determinants of mitochondrial content', *Human Molecular Genetics*, 16(12), pp. 1504-1514.

D'Erchia, A., Atlante, A., Gadaleta, G., Giulio, P., Chiara, M., Virgilio Caterina, D., Manzari, C., Francesca, M., Prazzoli, G.M., Picardi, E., Gissi, C., Horner, D., Aurelio, R., Sbisà, E., Tullo, A. and Pesole, G. (2014) *Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity*.

Da Silva, J.A. (1995) 'Sex hormones, glucocorticoids and autoimmunity: facts and hypotheses', *Annals of the Rheumatic Diseases*, 54(1), pp. 6-16.

Daley, D.O. and Whelan, J. (2005) 'Why genes persist in organelle genomes', *Genome Biology*, 6(5), pp. 110-110.

Damas, J., Carneiro, J., Amorim, A. and Pereira, F. (2014) 'MitoBreak: the mitochondrial DNA breakpoints database', *Nucleic Acids Res*, 42(Database issue), pp. D1261-8.

Danaei, G., Finucane, M.M., Lu, Y., Singh, G.M., Cowan, M.J., Paciorek, C.J., Lin, J.K., Farzadfar, F., Khang, Y.-H., Stevens, G.A., Rao, M., Ali, M.K., Riley, L.M., Robinson, C.A. and Ezzati, M. (2011) 'National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2·7 million participants', *The Lancet*, 378(9785), pp. 31-40.

Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.R., Iacono, W.G., Swaroop, A., Scott, L.J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G.R. and Fuchsberger, C. (2016) 'Next-generation genotype imputation service and methods', *Nat Genet*, 48(10), pp. 1284-1287.

Das, S., Morvan, F., Jourde, B., Meier, V., Kahle, P., Brebbia, P., Toussaint, G., Glass, David J. and Fornaro, M. (2015) 'ATP Citrate Lyase Improves Mitochondrial Function in Skeletal Muscle', *Cell Metabolism*, 21(6), pp. 868-876.

Dato, S., Passarino, G., Rose, G., Altomare, K., Bellizzi, D., Mari, V., Feraco, E., Franceschi, C. and De Benedictis, G. (2004) 'Association of the mitochondrial DNA haplogroup J with longevity is population specific', *Eur J Hum Genet*, 12(12), pp. 1080-1082.

Davalli, P., Mitic, T., Caporali, A., Lauriola, A. and D'Arca, D. (2016) 'ROS, Cell Senescence, and Novel Molecular Mechanisms in Aging and Age-Related Diseases', *Oxidative Medicine and Cellular Longevity*, 2016, p. 3565127.

Dayama, G., Emery, S.B., Kidd, J.M. and Mills, R.E. (2014) 'The genomic landscape of polymorphic human nuclear mitochondrial insertions', *Nucleic Acids Res*, 42.

de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S. and Voight, B.F. (2008) 'Practical aspects of imputation-driven meta-analysis of genome-wide association studies', *Hum Mol Genet*, 17(R2), pp. R122-8.

De Benedictis, G., Carrieri, G., Varcasia, O., Bonafè, M. and Franceschi, C. 908 (2000) 'Inherited variability of the mitochondrial genome and successful aging in humans' *Annals of the New York Academy of Sciences*. pp. 208-218.

de Bruin, J.P., Dorland, M., Spek, E.R., Posthuma, G., van Haaften, M., Looman, C.W.N. and te Velde, E.R. (2004) 'Age-Related Changes in the Ultrastructure of the Resting Follicle Pool in Human Ovaries1', *Biology of Reproduction*, 70(2), pp. 419-424.

de Lau, L.M. and Breteler, M.M. (2006) 'Epidemiology of Parkinson's disease', *Lancet Neurology*, 5(6), pp. 525-535.

Dean, S., Gould, M.K., Dewar, C.E. and Schnaufer, A.C. (2013) 'Single point mutations in ATP synthase compensate for mitochondrial genome loss in trypanosomes', *Proceedings of the National Academy of Sciences of the United States of America*, 110(36), pp. 14741-14746.

Delaneau, O., Marchini, J., Genomes Project, C. and Genomes Project, C. (2014) 'Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel', *Nat Commun*, 5, p. 3934.

Delaneau, O., Marchini, J. and Zagury, J.-F. (2012) 'A linear complexity phasing method for thousands of genomes', *Nat Meth*, 9(2), pp. 179-181.

Demarquoy, J. and Le Borgne, F. (2015) 'Crosstalk between mitochondria and peroxisomes', *World Journal of Biological Chemistry*, 6(4), pp. 301-309.

Diaz, F., Bayona-Bafaluy, M.P., Rana, M., Mora, M., Hao, H. and Moraes, C.T. (2002) 'Human mitochondrial DNA with large deletions repopulates organelles faster than full-length genomes under relaxed copy number control', *Nucleic Acids Research*, 30(21), pp. 4626-4633.

Dickinson, A., Yeung, K.Y., Donoghue, J., Baker, M.J., Kelly, R.D., McKenzie, M., Johns, T.G. and St. John, J.C. (2013) 'The regulation of mitochondrial DNA copy number in glioblastoma cells', *Cell Death and Differentiation*, 20(12), pp. 1644-1653.

Dimitrov, L., Lam, S.K. and Schekman, R. (2013) 'The Role of the Endoplasmic Reticulum in Peroxisome Biogenesis', *Cold Spring Harbor Perspectives in Biology*, 5(5), p. a013243.

Ding, J., Sidore, C., Butler, T.J., Wing, M.K., Qian, Y., Meirelles, O., Busonero, F., Tsoi, L.C., Maschio, A., Angius, A., Kang, H.M., Nagaraja, R., Cucca, F., Abecasis, G.R. and Schlessinger, D. (2015) 'Assessing Mitochondrial DNA Variation and Copy Number in Lymphocytes of ~2,000 Sardinians Using Tailored Sequencing Analysis Tools', *PLOS Genetics*, 11(7), p. e1005306.

Diroma, M.A., Calabrese, C. and Simone, D. (2014) 'Extraction and annotation of human mitochondrial genomes from 1000 Genomes Whole Exome Sequencing data', *BMC Genomics*, 15.

Doi, A., Park, I.H., Wen, B., Murakami, P. and Aryee, M.J. (2009) 'Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts', *Nat Genet*, 41.

Duan, Y., Li, F., Li, Y., Tang, Y., Kong, X., Feng, Z., Anthony, T.G., Watford, M., Hou, Y., Wu, G. and Yin, Y. (2016) 'The role of leucine and its metabolites in protein and energy metabolism', *Amino Acids*, 48(1), pp. 41-51.

Duan, Y., Zeng, L., Li, F., Wang, W., Li, Y., Guo, Q., Ji, Y., Tan, B. and Yin, Y. (2017) 'Effect of branched-chain amino acid ratio on the proliferation, differentiation, and expression levels of key regulators involved in protein metabolism of myocytes', *Nutrition*, 36, pp. 8-16.

Duchen, M.R. (2000) 'Mitochondria and calcium: From cell signalling to cell death', *Journal of Physiology*, 529(1), pp. 57-68.

Dupuis, L., Gonzalez de Aguilar, J.L., Oudart, H., de Tapia, M., Barbeito, L. and Loeffler, J.P. (2004) 'Mitochondria in amyotrophic lateral sclerosis: a trigger and a target', *Neurodegener Dis*, 1(6), pp. 245-54.

Durbin, R. (2014) 'Efficient haplotype matching and storage using the positional Burrows–Wheeler transform (PBWT)', *Bioinformatics*, 30(9), pp. 1266-1272.

Ebeling, P.R., Atley, L.M., Guthrie, J.R., Burger, H.G., Dennerstein, L., Hopper, J.L. and Wark, J.D. (1996) 'Bone turnover markers and bone density across the menopausal transition', *Journal of Clinical Endocrinology and Metabolism*, 81(9), pp. 3366-3371.

Ebos, J.M., Lee, C.R., Cruz-Munoz, W., Bjarnason, G.A., Christensen, J.G. and Kerbel, R.S. (2009) 'Accelerated metastasis after short-term treatment with a potent inhibitor of tumor angiogenesis', *Cancer Cell*, 15(3), pp. 232-9.

Edwards, Stacey L., Beesley, J., French, Juliet D. and Dunning, Alison M. (2013) 'Beyond GWASs: Illuminating the Dark Road from Association to Function', *The American Journal of Human Genetics*, 93(5), pp. 779-797.

Egerman, M.A. and Glass, D.J. (2014) 'Signaling pathways controlling skeletal muscle mass', *Critical Reviews in Biochemistry and Molecular Biology*, 49(1), pp. 59-68.

Egger, M., Davey Smith, G., Schneider, M. and Minder, C. (1997) 'Bias in meta-analysis detected by a simple, graphical test', *BMJ : British Medical Journal*, 315(7109), pp. 629-634.

Ekstrand, M.I., Falkenberg, M., Rantanen, A., Park, C.B., Gaspari, M., Hultenby, K., Rustin, P., Gustafsson, C.M. and Larsson, N.G. (2004) 'Mitochondrial transcription factor A regulates mtDNA copy number in mammals', *Human Molecular Genetics*, 13(9), pp. 935-944.

El-Hattab, A.W. and Scaglia, F. (2013) 'Mitochondrial DNA Depletion Syndromes: Review and Updates of Genetic Basis, Manifestations, and Therapeutic Options', *Neurotherapeutics*, 10(2), pp. 186-198.

Elliott, H.R., Samuels, D.C., Eden, J.A., Relton, C.L. and Chinnery, P.F. (2008) 'Pathogenic Mitochondrial DNA Mutations Are Common in the General Population', *American Journal of Human Genetics*, 83(2), pp. 254-260.

Elson, J.L., Andrews, R.M., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. and Howell, N. (2001) 'Analysis of European mtDNAs for recombination', *American Journal of Human Genetics*, 68(1), pp. 145-153.

Facucho-Oliveira, J.M., Alderson, J., Spikings, E.C., Egginton, S. and St. John, J.C. (2007) 'Mitochondrial DNA replication during differentiation of murine embryonic stem cells', *Journal of Cell Science*, 120(22), pp. 4025-4034.

Fairweather, D. and Rose, N.R. (2004) 'Women and Autoimmune Diseases', *Emerging Infectious Diseases*, 10(11), pp. 2005-2011.

Fan, A.X.C., R, R., Haghighi, M.M., Kohler, C., Xia, P., Hahn, S., Holzgreve, W. and Zhong, X.Y. (2008) 'Mitochondrial DNA content in paired normal and cancerous breast tissue samples from patients with breast cancer', in *J Cancer Res Clin Oncol*.

Fannin, S.W., Lesnefsky, E.J., Slabe, T.J., Hassan, M.O. and Hoppel, C.L. (1999) 'Aging Selectively Decreases Oxidative Capacity in Rat Heart Interfibrillar Mitochondria', *Archives of Biochemistry and Biophysics*, 372(2), pp. 399-407.

Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., Hatan, M., Carrasco-Alfonso, M.J., Mayer, D., Luckey, C.J., Patsopoulos, N.A., De Jager, P.L., Kuchroo, V.K.,

Epstein, C.B., Daly, M.J., Hafler, D.A. and Bernstein, B.E. (2015) 'Genetic and epigenetic fine mapping of causal autoimmune disease variants', *Nature*, 518(7539), pp. 337-343.

Farr, C.L., Wang, Y. and Kaguni, L.S. (1999) 'Functional interactions of mitochondrial dna polymerase and single- stranded DNA-binding protein. Template-primer DNA binding and initiation and elongation of DNA strand synthesis', *Journal of Biological Chemistry*, 274(21), pp. 14779-14785.

Fazzari, M.J. and Greally, J.M. (2004) 'Epigenomics: beyond CpG islands', *Nat Rev Genet*, 5(6), pp. 446-455.

Ferrini, R.L. and Barrett-Connor, E. (1998) 'Sex hormones and age: A cross-sectional study of testosterone and estradiol and their bioavailable fractions in community-dwelling men', *American Journal of Epidemiology*, 147(8), pp. 750-754.

Fink, G., Sumner, B.E.H., Rosie, R., Grace, O. and Quinn, J.P. (1996) 'Estrogen control of central neurotransmission: Effect on mood, mental state, and memory', *Cellular and Molecular Neurobiology*, 16(3), pp. 325-344.

Finkel, T. and Holbrook, N.J. (2000) 'Oxidants, oxidative stress and the biology of ageing', *Nature*, 408(6809), pp. 239-247.

Fischel-Ghodsian, N., Charlotte Bohlman, M., Prezant, T.R., Graham, J.M., Cederbaum, S.D. and Edwards, M.J. (1992) 'Deletion in blood mitochondrial DNA in kearns-sayre syndrome', *Pediatric Research*, 31(6), pp. 557-560.

Flannick, J. and Florez, J.C. (2016) 'Type 2 diabetes: genetic data sharing to advance complex disease research', *Nat Rev Genet*, 17(9), pp. 535-549.

Flindt, R. (2006) 'Life Span of Cells in Selected Organs of the Human Body', in Czeschlik, D. (ed.) *Amazing Numbers in Biology*. Germany: Springer, p. 204.

Frank, Steven A. (2012) 'Evolution: Mitochondrial Burden on Male Health', *Current Biology*, 22(18), pp. R797-R799.

Fraser, A., Macdonald-wallis, C., Tilling, K., Boyd, A., Golding, J., Davey smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., Ring, S., Nelson, S.M. and Lawlor, D.A. (2013) 'Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort', *International Journal of Epidemiology*, 42(1), pp. 97-110.

Friedman, J.R., Lackner, L.L., West, M., DiBenedetto, J.R., Nunnari, J. and Voeltz, G.K. (2011) 'ER tubules mark sites of mitochondrial division', *Science*, 334(6054), pp. 358-362.

Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D.J., Rivas, M.A., Perry, J.R.B., Sim, X., Blackwell, T.W., Robertson, N.R., Rayner, N.W., Cingolani, P., Locke, A.E., Tajes, J.F., Highland, H.M., Dupuis, J., Chines, P.S., Lindgren, C.M., Hartl, C., Jackson, A.U., Chen, H., Huyghe, J.R., Van De Bunt, M., Pearson, R.D., Kumar, A., Müller-Nurasyid, M., Grarup, N., Stringham, H.M., Gamazon, E.R., Lee, J., Chen, Y., Scott, R.A., Below, J.E., Chen, P., Huang, J., Go, M.J., Stitzel, M.L., Pasko, D., Parker, S.C.J., Varga, T.V., Green, T., Beer, N.L., Day-Williams, A.G., Ferreira, T., Fingerlin, T., Horikoshi, M., Hu, C., Huh, I., Ikram, M.K., Kim, B.J., Kim, Y., Kim, Y.J., Kwon, M.S., Lee, J., Lee, S., Lin, K.H., Maxwell, T.J., Nagai, Y., Wang, X., Welch, R.P., Yoon, J., Zhang, W., Barzilai, N., Voight, B.F., Han, B.G., Jenkinson, C.P., Kuulasmaa, T., Kuusisto, J., Manning, A., Ng, M.C.Y., Palmer, N.D., Balkau, B., Stančáková, A., Abboud, H.E., Boeing, H., Giedraitis, V., Prabhakaran, D., Gottesman, O., Scott, J., Carey, J., Kwan, P., Grant, G., Smith, J.D., Neale, B.M., Purcell, S., Butterworth, A.S., Howson, J.M.M., Lee, H.M., Lu, Y., Kwak, S.H., Zhao, W., Danesh, J., Lam, V.K.L., Park, K.S., Saleheen, D., et al. (2016) 'The genetic architecture of type 2 diabetes', *Nature*, 536(7614), pp. 41-47.

Fuke, S., Kubota-Sakashita, M., Kasahara, T., Shigeyoshi, Y. and Kato, T. (2011) 'Regional variation in mitochondrial DNA copy number in mouse brain', *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1807(3), pp. 270-274.

Fukuoh, A., Cannino, G., Gerards, M., Buckley, S., Kazancioglu, S., Scialo, F., Lihavainen, E., Ribeiro, A., Dufour, E. and Jacobs, H.T. (2014) 'Screen for mitochondrial DNA copy number maintenance genes reveals essential role for ATP synthase', *Molecular Systems Biology*, 10(6), pp. 734-734.

Gahan, M.E., Miller, F., Lewin, S.R., Cherry, C.L., Hoy, J.F., Mijch, A., Rosenfeldt, F. and Wesselingh, S.L. (2001) 'Quantification of mitochondrial DNA in peripheral blood mononuclear cells and subcutaneous fat using real-time polymerase chain reaction', *J Clin Virol*, 22(3), pp. 241-7.

Gaillard, S., Grasfeder, L.L., Haeffele, C.L., Lobenhofer, E.K., Chu, T.-M., Wolfinger, R., Kazmin, D., Koves, T.R., Muoio, D.M., Chang, C.-y. and McDonnell, D.P. (2006) 'Receptor-Selective Coactivators as Tools to Define the Biology of Specific Receptor-Coactivator Pairs', *Molecular Cell*, 24(5), pp. 797-803.

Gaillard, S.p., Dwyer, M.A. and McDonnell, D.P. (2007) 'Definition of the Molecular Basis for Estrogen Receptor-Related Receptor-α-Cofactor Interactions', *Molecular Endocrinology*, 21(1), pp. 62-76.

Gan, B., Hu, J., Jiang, S., Liu, Y., Sahin, E., Zhuang, L., Fletcher-Sananikone, E., Colla, S., Wang, Y.A., Chin, L. and DePinho, R.A. (2010) 'Lkb1 regulates quiescence and metabolic homeostasis of haematopoietic stem cells', *Nature*, 468(7324), pp. 701-704.

Garagnani, P., Pirazzini, C., Giuliani, C., Candela, M., Brigidi, P., Sevini, F., Luiselli, D., Bacalini, M.G., Salvioli, S., Capri, M., Monti, D., Mari, D., Collino, S., Delledonne, M., Descombes, P. and Franceschi, C. (2014) 'The Three Genetics (Nuclear DNA, Mitochondrial DNA, and Gut Microbiome) of Longevity in Humans Considered as Metaorganisms', *BioMed Research International*, 2014, p. 560340.

Garrido, N., Griparic, L., Jokitalo, E., Wartiovaara, J., van der Bliek, A.M. and Spelbrink, J.N. (2003) 'Composition and Dynamics of Human Mitochondrial Nucleoids', *Molecular Biology of the Cell*, 14(4), pp. 1583-1596.

Gaudet, A.D., Popovich, P.G. and Ramer, M.S. (2011) 'Wallerian degeneration: gaining perspective on inflammatory events after peripheral nerve injury', *Journal of Neuroinflammation*, 8, pp. 110-110.

Gemmell, N.J., Metcalf, V.J. and Allendorf, F.W. (2004) 'Mother's curse: the effect of mtDNA on individual fitness and population viability', *Trends in Ecology & Evolution*, 19(5), pp. 238-244.

Genome of the Netherlands, C. (2014) 'Whole-genome sequence variation, population structure and demographic history of the Dutch population', *Nat Genet*, 46(8), pp. 818-25.

Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A. and Abecasis, G.R. (2015) 'A global reference for human genetic variation', *Nature*, 526(7571), pp. 68-74.

Gervais, A.L., Marques, M. and Gaudreau, L. (2010) 'PCRTiler: automated design of tiled and specific PCR primer pairs', *Nucleic Acids Research*, 38(Web Server issue), pp. W308-W312.

Ghatak, S., Muthukumaran, R.B. and Nachimuthu, S.K. (2013) 'A Simple Method of Genomic DNA Extraction from Human Samples for PCR-RFLP Analysis', *Journal of Biomolecular Techniques : JBT*, 24(4), pp. 224-231.

Ghezzi, D., Marelli, C., Achilli, A., Goldwurm, S., Pezzoli, G., Barone, P., Pellecchia, M.T., Stanzione, P., Brusa, L., Bentivoglio, A.R., Bonuccelli, U., Petrozzi, L., Abbruzzese, G., Marchese, R., Cortelli, P., Grimaldi, D., Martinelli, P., Ferrarese, C., Garavaglia, B., Sangiorgi, S., Carelli, V., Torroni, A., Albanese, A. and Zeviani, M. (2005) 'Mitochondrial DNA haplogroup K is associated with a lower risk of parkinson's disease in Italians', *European Journal of Human Genetics*, 13(6), pp. 748-752.

Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic, J., Elling, U., Goodall, A.H., Labrune, Y., Lopez, L.M., Mägi, R., Meacham, S., Okada, Y., Pirastu, N., Sorice, R., Teumer, A., Voss, K., Zhang, W., Ramirez-Solis, R., Bis, J.C., Ellinghaus, D., Gögele, M., Hottenga, J.J., Langenberg, C., Kovacs, P., O'Reilly, P.F., Shin, S.Y., Esko, T., Hartiala, J., Kanoni, S., Murgia, F., Parsa, A., Stephens, J., Van Der Harst, P., Ellen Van Der Schoot, C., Allayee, H., Attwood, A., Balkau, B., Bastardot, F., Basu, S., Baumeister, S.E., Biino, G., Bomba, L., Bonnefond, A., Cambien, F., Chambers, J.C., Cucca, F., Dadamo, P., Davies, G., De Boer, R.A., De Geus, E.J.C., Döring, A., Elliott, P., Erdmann, J., Evans, D.M., Falchi, M., Feng, W., Folsom, A.R., Frazer, I.H., Gibson, Q.D., Glazer, N.L., Hammond, C., Hartikainen, A.L., Heckbert, S.R., Hengstenberg, C., Hersch, M., Illig, T., Loos, R.J.F., Jolley, J., Tee Khaw, K., Kühnel, B., Kyrtsonis, M.C., Lagou, V., Lloyd-Jones, H., Lumley, T., Mangino, M., Maschio, A., Mateo Leach, I., McKnight, B., Memari, Y., Mitchell, B.D., Montgomery, G.W., Nakamura, Y., Nauck, M., Navis, G., Nöthlings, U., Nolte, I.M., Porteous, D.J., Pouta, A., Pramstaller, P.P., Pullat, J., Ring, S.M., Rotter, J.I., Ruggiero, D., Ruokonen, A., Sala, C., Samani, N.J., Sambrook, J., Schlessinger, D., et al. (2011) 'New gene functions in megakaryopoiesis and platelet formation', *Nature*, 480(7376), pp. 201-208.

Gigli, I. and Bussmann, L.E. (2001) 'Exercise and ovarian steroid hormones', *Life Sciences*, 68(13), pp. 1505-1514.

Giles, R.E., Blanc, H., Cann, H.M. and Wallace, D.C. (1980) 'Maternal inheritance of human mitochondrial DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 77(11), pp. 6715-6719.

Gilkerson, R.W. (2009) 'Mitochondrial DNA nucleoids determine mitochondrial genetics and dysfunction', *The International Journal of Biochemistry & Cell Biology*, 41(10), pp. 1899-1906.

Giordano, C., Iommarini, L., Giordano, L., Maresca, A., Pisano, A., Valentino, M.L., Caporali, L., Liguori, R., Deceglie, S., Roberti, M., Fanelli, F., Fracasso, F., Ross-Cisneros, F.N., D'Adamo, P., Hudson, G., Pyle, A., Yu-Wai-Man, P., Chinnery, P.F., Zeviani, M., Salomao, S.R., Berezovsky, A., Belfort, R., Ventura, D.F., Moraes, M., Moraes Filho, M., Barboni, P., Sadun, F., De Negri, A., Sadun, A.A., Tancredi, A., Mancini, M., D'Amati, G., Loguercio Polosa, P., Cantatore, P. and Carelli, V. (2014) 'Efficient mitochondrial biogenesis drives incomplete penetrance in Leber's hereditary optic neuropathy', *Brain*, 137(2), pp. 335-353.

Gitschlag, Bryan L., Kirby, Cait S., Samuels, David C., Gangula, Rama D., Mallal, Simon A. and Patel, Maulik R. (2016) 'Homeostatic Responses Regulate Selfish Mitochondrial Genome Dynamics in C. elegans', *Cell Metabolism*, 24(1), pp. 91-103.

Gomes, A.P., Price, N.L., Ling, A.J.Y., Moslehi, J.J., Montgomery, M.K., Rajman, L., White, J.P., Teodoro, J.S., Wrann, C.D., Hubbard, B.P., Mercken, E.M., Palmeira, C.M., De Cabo, R., Rolo, A.P., Turner, N., Bell, E.L. and Sinclair, D.A. (2013) 'Declining NAD+ induces a pseudohypoxic state disrupting nuclear-mitochondrial communication during aging', *Cell*, 155(7), pp. 1624-1638.

Gomez-Duran, A., Pacheu-Grau, D., Lopez-Gallardo, E., Diez-Sanchez, C., Montoya, J., Lopez-Perez, M.J. and Ruiz-Pesini, E. (2010) 'Unmasking the causes of multifactorial disorders: OXPHOS differences between mitochondrial haplogroups', *Human Molecular Genetics*, 19(17), pp. 3343-53.

Gómez-Durán, A., Pacheu-Grau, D., Martínez-Romero, T., López-Gallardo, E., López-Pérez, M.J., Montoya, J. and Ruiz-Pesini, E. (2012) 'Oxidative phosphorylation differences between mitochondrial DNA haplogroups modify the risk of Leber's hereditary optic neuropathy', *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1822(8), pp. 1216-1222.

Gonzalez-Hunt, C.P., Rooney, J.P., Ryde, I.T., Anbalagan, C., Joglekar, R. and Meyer, J.N. (2016) 'PCR-based analysis of mitochondrial DNA copy number, mitochondrial DNA damage, and nuclear DNA damage', *Current protocols in toxicology / editorial board, Mahin D. Maines (editor-in-chief) ... [et al.]*, 67, pp. 20.11.1-20.11.25.

González-Vioque, E., Torres-Torronteras, J., Andreu, A.L. and Martí, R. (2011) 'Limited dCTP Availability Accounts for Mitochondrial DNA Depletion in Mitochondrial Neurogastrointestinal Encephalomyopathy (MNGIE)', *PLoS Genetics*, 7(3), p. e1002035.

Grady, J.P., Murphy, J.L., Blakely, E.L., Haller, R.G., Taylor, R.W., Turnbull, D.M. and Tuppen, H.A.L. (2014a) 'Accurate Measurement of Mitochondrial DNA Deletion Level and Copy Number Differences in Human Skeletal Muscle', *PLOS ONE*, 9(12), p. e114462.

Grady, J.P., Murphy, J.L., Blakely, E.L., Haller, R.G., Taylor, R.W., Turnbull, D.M. and Tuppen, H.A.L. (2014b) 'Accurate measurement of mitochondrial DNA deletion level and copy number differences in human skeletal muscle', *PLoS ONE*, 9(12).

Graf, E., Mahoney, J.R., Bryant, R.G. and Eaton, J.W. (1984) 'Iron-catalyzed hydroxyl radical formation. Stringent requirement for free iron coordination site', *Journal of Biological Chemistry*, 259(6), pp. 3620-3624.

Graham, A. (2015) 'Mitochondrial regulation of macrophage cholesterol homeostasis', *Free Radical Biology and Medicine*, 89, pp. 982-992.

Graham, B.H. 837 (2012) 'Diagnostic challenges of mitochondrial disorders: Complexities of two genomes' *Methods in Molecular Biology* [Article]. pp. 35-46.

Grand, R.S., Martienssen, R. and O'Sullivan, J.M. (2014) 'Potential roles for interactions between the mitochondrial and nuclear DNA throughout the cell cycle of Schizosaccharomyces pombe', *Mitochondrion*, 17(0), pp. 141-149.

Gray, L.R., Tompkins, S.C. and Taylor, E.B. (2014) 'Regulation of pyruvate metabolism and human disease', *Cellular and Molecular Life Sciences*, 71(14), pp. 2577-2604.

Gray, M.W. (2012) 'Mitochondrial evolution', *Cold Spring Harbor Perspectives in Biology*, 4(9).

Groot, G.S.P. and Kroon, A.M. (1979) 'Mitochondrial DNA from various organisms does not contain internally methylated cytosine in -CCGG- sequences', *BBA Section Nucleic Acids And Protein Synthesis*, 564(2), pp. 355-357.

Grünewald, A., Lax, N.Z., Rocha, M.C., Reeve, A.K., Hepplewhite, P.D., A. Rygiel, K., Taylor, R.W. and Turnbull, D.M. (2014) 'Quantitative quadruple-label immunofluorescence of mitochondrial and cytoplasmic proteins in single neurons from human midbrain tissue', *Journal of Neuroscience Methods*, 232(100), pp. 143-149.

Gu, F., Chauhan, V., Kaur, K., Brown, W.T., LaFauci, G., Wegiel, J. and Chauhan, A. (2013) 'Alterations in mitochondrial DNA copy number and the activities of electron transport chain complexes and pyruvate dehydrogenase in the frontal cortex from subjects with autism', *Transl Psychiatry*, 3, p. e299.

Guo, X., Fan, W., Bian, X. and Ma, D. (2014) 'Upregulation of the Kank1 gene-induced brain glioma apoptosis and blockade of the cell cycle in G0/G1 phase', *Int J Oncol*, 44(3), pp. 797-804.

Guo, Y., Li, J., Li, C.-I., Shyr, Y. and Samuels, D.C. (2013) 'MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis', *Bioinformatics*, 29(9), pp. 1210-1211.

Gurung, P., Lukens, J.R. and Kanneganti, T.-D. 'Mitochondria: diversity in the regulation of the NLRP3 inflammasome', *Trends in Molecular Medicine*, 21(3), pp. 193-201.

Guyatt, A.L., Burrows, K., Guthrie, P.A.I., Ring, S., McArdle, W., Day, I.N.M., Ascione, R., Lawlor, D.A., Gaunt, T.R. and Rodriguez, S. (2017) 'Cardiometabolic phenotypes and mitochondrial DNA copy number in two cohorts of UK women', *Mitochondrion*.

Gwak, M.S., Choi, S.J., Kim, J.A., Ko, J.S., Kim, T.H., Lee, S.M., Park, J.-A. and Kim, M.H. (2007) 'Effects of Gender on White Blood Cell Populations and Neutrophil-Lymphocyte Ratio Following Gastrectomy in Patients with Stomach Cancer', *Journal of Korean Medical Science*, 22(Suppl), pp. S104-S108.

Haas, R.H., Parikh, S., Falk, M.J., Saneto, R.P., Wolf, N.I., Darin, N., Wong, L.-J., Cohen, B.H. and Naviaux, R.K. (2008) 'The In-Depth Evaluation of Suspected Mitochondrial Disease: The Mitochondrial Medicine Society's Committee on Diagnosis', *Molecular genetics and metabolism*, 94(1), pp. 16-37.

Haemmerle, G., Moustafa, T., Woelkart, G., Büttner, S., Schmidt, A., Van De Weijer, T., Hesselink, M., Jaeger, D., Kienesberger, P.C., Zierler, K., Schreiber, R., Eichmann, T., Kolb, D., Kotzbeck, P., Schweiger, M., Kumari, M., Eder, S., Schoiswohl, G., Wongsiriroj, N., Pollak, N.M., Radner, F.P.W., Preiss-Landl, K., Kolbe, T., Rülicke, T., Pieske, B., Trauner, M., Lass, A., Zimmermann, R., Hoefler, G., Cinti, S., Kershaw, E.E., Schrauwen, P., Madeo, F., Mayer, B. and Zechner, R. (2011) 'ATGL-mediated fat catabolism regulates cardiac mitochondrial function via PPAR-Î ± and PGC-1', *Nature Medicine*, 17(9), pp. 1076-1085.

Halestrap, A.P. (1975) 'The mitochondrial pyruvate carrier. Kinetics and specificity for substrates and inhibitors', *Biochemical Journal*, 148(1), pp. 85-96.

Hallberg, B.M. and Larsson, N.-G. (2011) 'TFAM forces mtDNA to make a U-turn', *Nat Struct Mol Biol*, 18(11), pp. 1179-1181.

Harman, D. (1956) 'Aging: A Theory Based on Free Radical and Radiation Chemistry', *Journal of Gerontology*, 11(3), pp. 298-300.

Harrington, J.S., Choi, A.M.K. and Nakahira, K. (2017) 'Mitochondrial DNA in Sepsis', *Current Opinion in Critical Care*, 23(4), pp. 284-290.

Harvey, A., Gibson, T., Lonergan, T. and Brenner, C. (2011) 'Dynamic regulation of mitochondrial function in preimplantation embryos and embryonic stem cells', *Mitochondrion*, 11(5), pp. 829-838.

Hawley, S.A., Pan, D.A., Mustard, K.J., Ross, L., Bain, J., Edelman, A.M., Frenguelli, B.G. and Hardie, D.G. (2005) 'Calmodulin-dependent protein kinase kinase-β is an alternative upstream kinase for AMP-activated protein kinase', *Cell Metabolism*, 2(1), pp. 9-19.

Hay, M., . (2004) 'Eupedia.com Genetics' December 2004. July 2017. Available at: http://www.eupedia.com/europe/origins_haplogroups_europe.shtml.

Hazkani-Covo, E. and Covo, S. (2008) 'Numt-mediated double-strand break repair mitigates deletions during primate genome evolution', *PLoS Genet*, 4.

Hazkani-Covo, E. and Martin, W.F. (2017) 'Quantifying the number of independent organelle DNA insertions in genome evolution and human health', *Genome Biol Evol*.

Hazkani-Covo, E., Zeller, R.M. and Martin, W. (2010) 'Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes', *PLoS Genet*, 6.

He, L., Chinnery, P.F., Durham, S.E., Blakely, E.L., Wardell, T.M., Borthwick, G.M., Taylor, R.W. and Turnbull, D.M. (2002) 'Detection and quantification of mitochondrial DNA deletions in individual cells by real-time PCR', *Nucleic acids research*, 30(14).

He, Y., Tang, J., Li, Z., Li, H., Liao, Y., Tang, Y., Tan, L., Chen, J., Xia, K. and Chen, X. (2014) 'Leukocyte Mitochondrial DNA Copy Number in Blood Is Not Associated with Major Depressive Disorder in Young Adults', *PLOS ONE*, 9(5), p. e96869.

Heddi, A., Stepien, G., Benke, P.J. and Wallace, D.C. (1999) 'Coordinate induction of energy gene expression in tissues of mitochondrial disease patients', *J Biol Chem*, 274(33), pp. 22968-76.

Hegde, M.L., Mantha, A.K., Hazra, T.K., Bhakat, K.K., Mitra, S. and Szczesny, B. (2012) 'Oxidative genome damage and its repair: Implications in aging and neurodegenerative diseases', *Mechanisms of Ageing and Development*, 133(4), pp. 157-168.

Heid, C.A., Stevens, J., Livak, K.J. and Williams, P.M. (1996) 'Real time quantitative PCR', *Genome Research*, 6(10), pp. 986-994.

Hendrickson, S.L., Hutcheson, H.B., Ruiz-Pesini, E., Poole, J.C., Lautenberger, J., Sezgin, E., Kingsley, L., Goedert, J.J., Vlahov, D., Donfield, S., Wallace, D.C. and O'Brien, S.J. (2008) 'Mitochondrial DNA haplogroups influence AIDS progression', *AIDS*, 22(18), pp. 2429-2439.

Henrich, T.J., Gallien, S., Li, J.Z., Pereyra, F. and Kuritzkes, D.R. (2012) 'Low-Level Detection and Quantitation of Cellular HIV-1 DNA and 2-LTR Circles Using Droplet Digital PCR', *Journal of virological methods*, 186(1-2), pp. 68-72.

Hernández-Ríos, R., Hernández-Estrada, S., Cruz-Robles, D., Hernández-Lobato, S., Villalobos-Martín, M., Johnson, R.J., Rodríguez-Castellanos, F., Salazar, J.,

García-Arroyo, F., Sánchez-Lozada, L.G. and Madero, M. (2013) 'Low fructose and low salt diets increase mitochondrial DNA in white blood cells of overweight subjects', *Experimental and Clinical Endocrinology and Diabetes*, 121(9), pp. 535-538.

Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., Vilella, A.J., Searle, S.M.J., Amode, R., Brent, S., Spooner, W., Kulesha, E., Yates, A. and Flicek, P. (2016) 'Ensembl comparative genomics resources', *Database*, 2016, pp. bav096-bav096.

Higuchi, M. (2007) 'Regulation of mitochondrial DNA content and cancer', *Mitochondrion*, 7.

Hill-Burns, E.M., Debelius, J.W., Morton, J.T., Wissemann, W.T., Lewis, M.R., Wallen, Z.D., Peddada, S.D., Factor, S.A., Molho, E., Zabetian, C.P., Knight, R. and Payami, H. (2017) 'Parkinson's disease and Parkinson's disease medications have distinct signatures of the gut microbiome', *Movement Disorders*, 32(5), pp. 739-749.

Hindorff LA., MacArthur J., Morales J., Junkins HA., Hall PN., Klemm AK. and Manolio TA. (2017) 'A Catalog of Published Genome-Wide Association Studies'. 8th June 2017. European Bioinformatics Institute. Available at: http://www.genome.gov/gwastudies.

Hoaglin, D.C. (2016) 'Misunderstandings about Q and 'Cochran's Q test' in meta-analysis', *Statistics in Medicine*, 35(4), pp. 485-495.

Holland, P.M., Abramson, R.D., Watson, R. and Gelfand, D.H. (1991) 'Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase', *Proceedings of the National Academy of Sciences of the United States of America*, 88(16), pp. 7276-7280.

Hollenbeck, P.J. and Saxton, W.M. (2005) 'The axonal transport of mitochondria', *Journal of Cell Science*, 118(23), pp. 5411-5419.

Holt, I.J., Harding, A.E., Petty, R.K.H. and Morgan-Hughes, J.A. (1990) 'A new mitochondrial disease associated with mitochondrial DNA heteroplasmy', *American Journal of Human Genetics*, 46(3), pp. 428-433.

Holt, I.J., Lorimer, H.E. and Jacobs, H.T. (2000) 'Coupled Leading- and Lagging-Strand Synthesis of Mammalian Mitochondrial DNA', *Cell*, 100(5), pp. 515-524.

Hon, C.-C., Ramilowski, J.A., Harshbarger, J., Bertin, N., Rackham, O.J.L., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T.M., Severin, J., Lizio, M., Kawaji, H., Kasukawa, T., Itoh, M., Burroughs, A.M., Noma, S., Djebali, S., Alam, T., Medvedeva, Y.A., Testa, A.C., Lipovich, L., Yip, C.-W., Abugessaisa, I., Mendez, M., Hasegawa, A., Tang, D., Lassmann, T., Heutink, P., Babina, M., Wells, C.A., Kojima, S., Nakamura, Y., Suzuki, H., Daub, C.O., de Hoon, M.J.L., Arner, E., Hayashizaki, Y., Carninci, P. and Forrest, A.R.R. (2017) 'An atlas of human long non-coding RNAs with accurate 5′ ends', *Nature*, 543(7644), pp. 199-204.

Hoppins, S., Collins, S.R., Cassidy-Stone, A., Hummel, E., DeVay, R.M., Lackner, L.L., Westermann, B., Schuldiner, M., Weissman, J.S. and Nunnari, J. (2011) 'A mitochondrial-focused genetic interaction map reveals a scaffold-like complex required for inner membrane organization in mitochondria', *The Journal of Cell Biology*, 195(2), pp. 323-340.

Hori, A., Yoshida, M. and Ling, F. (2011) 'Mitochondrial fusion increases the mitochondrial DNA copy number in budding yeast', *Genes to Cells*, 16(5), pp. 527-544.

Horvath, R., Hudson, G., Ferrari, G., Futterer, N., Ahola, S., Lamantea, E., Prokisch, H., Lochmuller, H., McFarland, R., Ramesh, V., Klopstock, T., Freisinger, P., Salvi, F., Mayr, J.A., Santer, R., Tesarova, M., Zeman, J., Udd, B., Taylor, R.W., Turnbull, D., Hanna, M., Fialho, D., Suomalainen, A., Zeviani, M. and Chinnery, P.F. (2006) 'Phenotypic spectrum associated with mutations of the mitochondrial polymerase gamma gene', *Brain*, 129(Pt 7), pp. 1674-84.

Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K. and Kelsey, K.T. (2012) 'DNA methylation arrays as surrogate measures of cell mixture distribution', *BMC Bioinformatics*, 13(1), pp. 1-16.

Howie, B.N., Donnelly, P. and Marchini, J. (2009) 'A flexible and accurate genotype imputation method for the next generation of genome-wide association studies', *PLoS Genet*, 5(6), p. e1000529.

Huang, J., Howie, B., McCarthy, S., Memari, Y., Walter, K., Min, J.L., Danecek, P., Malerba, G., Trabetti, E., Zheng, H.F., Consortium, U.K., Gambaro, G., Richards, J.B., Durbin, R., Timpson, N.J., Marchini, J. and Soranzo, N. (2015) 'Improved

imputation of low-frequency and rare variants using the UK10K haplotype reference panel', *Nat Commun*, 6, p. 8111.

Hudson, G., Carelli, V., Horvath, R., Zeviani, M., Smeets, H.J. and Chinnery, P.F. (2007a) 'X-inactivation patterns in females harboring mtDNA mutations that cause Leber hereditary optic neuropathy', *Molecular Vision*, 13, pp. 2339-2343.

Hudson, G., Carelli, V., Spruijt, L., Gerards, M., Mowbray, C., Achilli, A., Pyle, A., Elson, J., Howell, N., La Morgia, C., Valentino, M.L., Huoponen, K., Savontaus, M.-L., Nikoskelainen, E., Sadun, A.A., Salomao, S.R., Belfort Jr, R., Griffiths, P., Man, P.Y.W., de Coo, R.F.M., Horvath, R., Zeviani, M., Smeets, H.J.T., Torroni, A. and Chinnery, P.F. (2007b) 'Clinical Expression of Leber Hereditary Optic Neuropathy Is Affected by the Mitochondrial DNA–Haplogroup Background', *The American Journal of Human Genetics*, 81(2), pp. 228-233.

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

Hudson, G., Nalls, M., Evans, J.R., Breen, D.P., Winder-Rhodes, S., Morrison, K.E., Morris, H.R., Williams-Gray, C.H., Barker, R.A., Singleton, A.B., Hardy, J., Wood, N.E., Burn, D.J. and Chinnery, P.F. (2013a) 'Two-stage association study and meta-analysis of mitochondrial DNA variants in Parkinson disease', *Neurology*, 80(22), pp. 2042-2048.

Hudson, G., Panoutsopoulou, K., Wilson, I., Southam, L., Rayner, N.W., Arden, N., Birrell, F., Carluke, I., Carr, A., Chapman, K., Deloukas, P., Doherty, M., McCaskie, A., Ollier, W.E.R., Ralston, S.H., Reed, M.R., Spector, T.D., Valdes, A.M., Wallis, G.A., Wilkinson, J.M., Zeggini, E., Samuels, D.C., Loughlin, J. and Chinnery, P.F. (2013b) 'No evidence of an association between mitochondrial DNA variants and osteoarthritis in 7393 cases and 5122 controls', *Annals of the Rheumatic Diseases*, 72(1), pp. 136-139.

Hulgan, T., Haubrich, R., Riddler, S.A., Tebas, P., Ritchie, M.D., McComsey, G.A., Haas, D.W. and Canter, J.A. (2011) 'European mitochondrial DNA haplogroups and metabolic changes during antiretroviral therapy in AIDS Clinical Trials Group Study A5142', *AIDS*, 25(1), pp. 37-47.

Hurtado-Roca, Y., Ledesma, M., Gonzalez-Lazaro, M., Moreno-Loshuertos, R., Fernandez-Silva, P., Enriquez, J.A. and Laclaustra, M. (2016) 'Adjusting MtDNA Quantification in Whole Blood for Peripheral Blood Platelet and Leukocyte Counts', *PLoS One*, 11(10), p. e0163770.

Hutchison, C.A., Newbold, J.E., Potter, S.S. and Edgell, M.H. (1974) 'Maternal inheritance of mammalian mitochondrial DNA', *Nature*, 251(5475), pp. 536-538.

Hyslop, L.A., Blakeley, P., Craven, L., Richardson, J., Fogarty, N.M.E., Fragouli, E., Lamb, M., Wamaitha, S.E., Prathalingam, N., Zhang, Q., O'Keefe, H., Takeda, Y., Arizzi, L., Alfarawati, S., Tuppen, H.A., Irving, L., Kalleas, D., Choudhary, M., Wells, D., Murdoch, A.P., Turnbull, D.M., Niakan, K.K. and Herbert, M. (2016) 'Towards clinical application of pronuclear transfer to prevent mitochondrial DNA disease', *Nature*, 534(7607), pp. 383-386.

IBM Corporation (2013) *IBM SPSS Statistics for Windows* (Version 22) [Computer program]. IBM Corporation. Available at: http://www.ibm.com/us/en/.

Improta-Brears, T., Whorton, A.R., Codazzi, F., York, J.D., Meyer, T. and McDonnell, D.P. (1999) 'Estrogen-induced activation of mitogen-activated protein kinase requires mobilization of intracellular calcium', *Proceedings of the National Academy of Sciences of the United States of America*, 96(8), pp. 4686-4691.

Indiana University (2014) *Obesity, Type 2 Diabetes, and Fructose*. Available at: http://www.indiana.edu/~oso/Fructose/MolBiol.html (Accessed: 15/05/2017).

Ingman, M., Kaessmann, H., Pääbo, S. and Gyllensten, U. (2000) 'Mitochondrial genome variation and the origin of modem humans', *Nature*, 408(6813), pp. 708-713.

International HapMap, C. (2003) 'The International HapMap Project', *Nature*, 426(6968), pp. 789-96.

International HapMap, C. (2005) 'A haplotype map of the human genome', *Nature*, 437(7063), pp. 1299-320.

International Multiple Sclerosis Genetics, C., Wellcome Trust Case Control, C., Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E., Edkins, S., Gray, E., Booth, D.R., Potter, S.C., Goris, A., Band, G., Oturai, A.B., Strange, A., Saarela, J.,

Bellenguez, C., Fontaine, B., Gillman, M., Hemmer, B., Gwilliam, R., Zipp, F., Jayakumar, A., Martin, R., Leslie, S., Hawkins, S., Giannoulatou, E., D'Alfonso, S., Blackburn, H., Martinelli Boneschi, F., Liddle, J., Harbo, H.F., Perez, M.L., Spurkland, A., Waller, M.J., Mycko, M.P., Ricketts, M., Comabella, M., Hammond, N., Kockum, I., McCann, O.T., Ban, M., Whittaker, P., Kemppinen, A., Weston, P., Hawkins, C., Widaa, S., Zajicek, J., Dronov, S., Robertson, N., Bumpstead, S.J., Barcellos, L.F., Ravindrarajah, R., Abraham, R., Alfredsson, L., Ardlie, K., Aubin, C., Baker, A., Baker, K., Baranzini, S.E., Bergamaschi, L., Bergamaschi, R., Bernstein, A., Berthele, A., Boggild, M., Bradfield, J.P., Brassat, D., Broadley, S.A., Buck, D., Butzkueven, H., Capra, R., Carroll, W.M., Cavalla, P., Celius, E.G., Cepok, S., Chiavacci, R., Clerget-Darpoux, F., Clysters, K., Comi, G., Cossburn, M., Cournu-Rebeix, I., Cox, M.B., Cozen, W., Cree, B.A., Cross, A.H., Cusi, D., Daly, M.J., Davis, E., de Bakker, P.I., Debouverie, M., D'Hooghe M, B., Dixon, K., Dobosi, R., Dubois, B., Ellinghaus, D., et al. (2011) 'Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis', *Nature*, 476(7359), pp. 214-9.

J Nicholls, T. and Minczuk, M. (2014) *In D-loop: 40years of mitochondrial 7S DNA*.

Jacobs, L.J.A.M., de Wert, G., Geraedts, J.P.M., de Coo, I.F.M. and Smeets, H.J.M. (2006) 'The transmission of OXPHOS disease and methods to prevent this', *Human Reproduction Update*, 12(2), pp. 119-136.

Jaenisch, R. and Bird, A. 'Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals', *Nat Genet*.

Jaffe, A.E. and Irizarry, R.A. (2014) 'Accounting for cellular heterogeneity is critical in epigenome-wide association studies', *Genome Biol*, 15(2), p. R31.

Jajoo, R., Jung, Y., Huh, D., Viana, M.P., Rafelski, S.M., Springer, M. and Paulsson, J. (2016) 'Accurate concentration control of mitochondria and nucleoids', *Science*, 351(6269), pp. 169-172.

Jemt, E., Persson, Ö., Shi, Y., Mehmedovic, M., Uhler, J.P., Dávila López, M., Freyer, C., Gustafsson, C.M., Samuelsson, T. and Falkenberg, M. (2015) 'Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element', *Nucleic Acids Research*, 43(19), pp. 9262-9275.

Jensen, E.V. and DeSombre, E.R. (1973) 'Estrogen-Receptor Interaction', *Estrogenic hormones effect transformation of specific receptor proteins to a biochemically functional form*, 182(4108), pp. 126-134.

Jenuth, J.P., Peterson, A.C., Fu, K. and Shoubridge, E.A. (1996) 'Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA', *Nat Genet*, 14(2), pp. 146-151.

Jia, H., Guo, Y., Zhao, W. and Wang, K. (2014) 'Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer', *Scientific Reports*, 4, p. 5737.

Jia, J.J., Tian, Y.B., Cao, Z.H., Tao, L.L., Zhang, X., Gao, S.Z., Ge, C.R., Lin, Q.Y. and Jois, M. (2010) 'The polymorphisms of UCP1 genes associated with fat metabolism, obesity and diabetes', *Mol Biol Rep*, 37(3), pp. 1513-22.

Jiang, W.W., Masayesva, B., Zahurak, M., Carvalho, A.L., Rosenbaum, E., Mambo, E., Zhou, S., Minhas, K., Benoit, N., Westra, W.H., Alberg, A., Sidransky, D., Koch, W. and Califano, J. (2005) 'Increased mitochondrial DNA content in saliva associated with head and neck cancer', *Clin Cancer Res*, 11.

Jiang, W.W., Rosenbaum, E., Mambo, E., Zahurak, M., Masayesva, B., Carvalho, A.L., Zhou, S., Westra, W.H., Alberg, A.J., Sidransky, D., Koch, W. and Califano, J.A. (2006) 'Decreased mitochondrial DNA content in posttreatment salivary rinses from head and neck cancer patients', *Clin Cancer Res*, 12.

Jin, R.U. and Mills, J.C. (2014) 'RAB26 coordinates lysosome traffic and mitochondrial localization', *Journal of Cell Science*, 127(5), pp. 1018-1032.

Jitrapakdee, S., St Maurice, M., Rayment, I., Cleland, W.W., Wallace, J.C. and Attwood, P.V. (2008) 'Structure, mechanism and regulation of pyruvate carboxylase', *Biochemical Journal*, 413(3), pp. 369-387.

Johnson, M.J., Wallace, D.C., Ferris, S.D., Rattazzi, M.C. and Cavalli-Sforza, L.L. (1983) 'Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns', *J Mol Evol*, 19(3-4), pp. 255-71.

Johnston, I.G. and Williams, B.P. (2016) 'Evolutionary Inference across Eukaryotes Identifies Specific Pressures Favoring Mitochondrial Gene Retention', *Cell Syst*, 2(2), pp. 101-11.

Jones, C.N., Wilkinson, K.A., Hung, K.T., Weeks, K.M. and Spremulli, L.L. (2008) 'Lack of secondary structure characterizes the 5′ ends of mammalian mitochondrial mRNAs', *RNA*, 14(5), pp. 862-871.

Jørgensen, W., Rud, K.A., Mortensen, O.H., Frandsen, L., Grunnet, N. and Quistorff, B. (2017) 'Your mitochondria are what you eat: a high-fat or a high-sucrose diet eliminates metabolic flexibility in isolated mitochondria from rat skeletal muscle', *Physiological Reports*, 5(6).

Jornayvaz, F.R. and Shulman, G.I. (2010) 'Regulation of mitochondrial biogenesis', *Essays in biochemistry*, 47, p. 10.1042/bse0470069.

Ju, Y.S., Tubio, J.M., Mifsud, W., Fu, B., Davies, H.R., Ramakrishna, M., Li, Y., Yates, L., Gundem, G., Tarpey, P.S., Behjati, S., Papaemmanuil, E., Martin, S., Fullam, A., Gerstung, M., Group, I.P.C.W., Group, I.B.C.W., Group, I.B.C.W., Nangalia, J., Green, A.R., Caldas, C., Borg, A., Tutt, A., Lee, M.T., van't Veer, L.J., Tan, B.K., Aparicio, S., Span, P.N., Martens, J.W., Knappskog, S., Vincent-Salomon, A., Borresen-Dale, A.L., Eyfjord, J.E., Myklebost, O., Flanagan, A.M., Foster, C., Neal, D.E., Cooper, C., Eeles, R., Bova, S.G., Lakhani, S.R., Desmedt, C., Thomas, G., Richardson, A.L., Purdie, C.A., Thompson, A.M., McDermott, U., Yang, F., Nik-Zainal, S., Campbell, P.J. and Stratton, M.R. (2015) 'Frequent somatic transfer of mitochondrial DNA into the nuclear genome of human cancer cells', *Genome Res*, 25(6), pp. 814-24.

Kang, D., Kim, S.H. and Hamasaki, N. (2007) 'Mitochondrial transcription factor A (TFAM): Roles in maintenance of mtDNA and cellular functions', *Mitochondrion*, 7(1–2), pp. 39-44.

Kannarkat, G.T., Boss, J.M. and Tansey, M.G. (2013) 'The Role of Innate and Adaptive Immunity in Parkinson's Disease', *Journal of Parkinson's disease*, 3(4), pp. 493-514.

Kasashima, K., Sumitani, M. and Endo, H. (2011) 'Human mitochondrial transcription factor A is required for the segregation of mitochondrial DNA in cultured cells', *Experimental Cell Research*, 317(2), pp. 210-220.

Kaufman, B.A., Durisic, N., Mativetsky, J.M., Costantino, S., Hancock, M.A., Grutter, P. and Shoubridge, E.A. (2007) 'The Mitochondrial Transcription Factor TFAM Coordinates the Assembly of Multiple DNA Molecules into Nucleoid-like Structures', *Molecular Biology of the Cell*, 18(9), pp. 3225-3236.

Kaukonen, J., Juselius, J.K., Tiranti, V., Kyttala, A., Zeviani, M., Comi, G.P., Keranen, S., Peltonen, L. and Suomalainen, A. (2000) 'Role of adenine nucleotide translocator 1 in mtDNA maintenance', *Science*, 289(5480), pp. 782-785.

Keeney, P.M. and Bennett, J.P. (2010) 'ALS spinal neurons show varied and reduced mtDNA gene copy numbers and increased mtDNA gene deletions', *Molecular Neurodegeneration*, 5, pp. 21-21.

Kelly, D.P. and Scarpulla, R.C. (2004) 'Transcriptional regulatory circuits controlling mitochondrial biogenesis and function', *Genes and Development*, 18(4), pp. 357-368.

Kelly, R.D., Rodda, A.E., Dickinson, A., Mahmud, A., Nefzger, C.M., Lee, W., Forsythe, J.S., Polo, J.M., Trounce, I.A., McKenzie, M., Nisbet, D.R. and St John, J.C. (2013a) 'Mitochondrial DNA haplotypes define gene expression patterns in pluripotent and differentiating embryonic stem cells', *Stem Cells*, 31(4), pp. 703-16.

Kelly, R.D., Sumer, H., McKenzie, M., Facucho-Oliveira, J., Trounce, I.A., Verma, P.J. and St John, J.C. (2013b) 'The effects of nuclear reprogramming on mitochondrial DNA replication', *Stem Cell Rev*, 9(1), pp. 1-15.

Kelly, R.D.W., Mahmud, A., McKenzie, M., Trounce, I.A. and St John, J.C. (2012) 'Mitochondrial DNA copy number is regulated in a tissue specific manner by DNA methylation of the nuclear-encoded DNA polymerase gamma A', *Nucleic Acids Research*, 40(20), pp. 10124-10138.

Kemper, M.F., Stirone, C., Krause, D.N., Duckles, S.P. and Procaccio, V. (2014) 'Genomic and non-genomic regulation of PGC1 isoforms by estrogen to increase cerebral vascular mitochondrial biogenesis and reactive oxygen species protection', *European Journal of Pharmacology*, 723(1), pp. 322-329.

Kenmochi, N., Suzuki, T., Uechi, T., Magoori, M., Kuniba, M., Higa, S., Watanabe, K. and Tanaka, T. (2001) 'The human mitochondrial ribosomal protein genes: Mapping of 54 genes to the chromosomes and implications for human disorders', *Genomics*, 77(1-2), pp. 65-70.

Khalifa, A.R.M., Abdel‐Rahman, E.A., Mahmoud, A.M., Ali, M.H., Noureldin, M., Saber, S.H., Mohsen, M. and Ali, S.S. (2017) 'Sex‐specific differences in mitochondria biogenesis, morphology, respiratory function, and ROS homeostasis in young mouse heart and brain', *Physiological Reports*, 5(6), p. e13125.

Khor, B., Gardet, A. and Xavier, R.J. (2011) 'Genetics and pathogenesis of inflammatory bowel disease', *Nature*, 474(7351), pp. 307-317.

Kilbaugh, T.J., Lvova, M., Karlsson, M., Zhang, Z., Leipzig, J., Wallace, D.C. and Margulies, S.S. (2015) 'Peripheral Blood Mitochondrial DNA as a Biomarker of Cerebral Mitochondrial Dysfunction following Traumatic Brain Injury in a Porcine Model', *PLOS ONE*, 10(6), p. e0130927.

Kim, H.J., Magranè, J., Starkov, A.A. and Manfredi, G. (2012) 'The mitochondrial calcium regulator cyclophilin D is an essential component of oestrogen-mediated neuroprotection in amyotrophic lateral sclerosis', *Brain*, 135(9), pp. 2865-2874.

Kim, J., Kollhoff, A., Bergmann, A. and Stubbs, L. (2003) 'Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3', *Human Molecular Genetics*, 12(3), pp. 233-245.

Kim, Y. and Park, Y. (2015) 'Conjugated Linoleic Acid (CLA) Stimulates Mitochondrial Biogenesis Signaling by the Upregulation of PPARγ Coactivator 1α (PGC-1α) in C2C12 Cells', *Lipids*, 50(4), pp. 329-338.

Kitchen, R.R., Kubista, M. and Tichopad, A. (2010) 'Statistical aspects of quantitative real-time PCR experiment design', *Methods*, 50(4), pp. 231-236.

Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G. and Kronenberg, F. (2011) 'HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups', *Human Mutation*, 32(1), pp. 25-32.

Knez, J., Winckelmans, E., Plusquin, M., Thijs, L., Cauwenberghs, N., Gu, Y., Staessen, J.A., Nawrot, T.S. and Kuznetsova, T. (2016) 'Correlates of Peripheral Blood Mitochondrial DNA Content in a General Population', *American Journal of Epidemiology*, 183(2), pp. 138-146.

Kofler, B., Mueller, E.E., Eder, W., Stanger, O., Maier, R., Weger, M., Haas, A., Winker, R., Schmut, O., Paulweber, B., Iglseder, B., Renner, W., Wiesbauer, M., Aigner, I., Santic, D., Zimmermann, F.A., Mayr, J.A. and Sperl, W. (2009a) 'Mitochondrial DNA haplogroup T is associated with coronary artery disease and diabetic retinopathy: a case control study', *BMC Medical Genetics*, 10, pp. 35-35.

Kofler, B., Mueller, E.E., Eder, W., Stanger, O., Maier, R., Weger, M., Haas, A., Winker, R., Schmut, O., Paulweber, B., Iglseder, B., Renner, W., Wiesbauer, M., Aigner, I., Santic, D., Zimmermann, F.A., Mayr, J.A. and Sperl, W. (2009b) 'Mitochondrial DNA haplogroup T is associated with coronary artery disease and diabetic retinopathy: a case control study', *BMC Med Genet*, 10, p. 35.

Konokhova, Y., Spendiff, S., Jagoe, R.T., Aare, S., Kapchinsky, S., MacMillan, N.J., Rozakis, P., Picard, M., Aubertin-Leheudre, M., Pion, C.H., Bourbeau, J., Hepple, R.T. and Taivassalo, T. (2016) 'Failed upregulation of TFAM protein and mitochondrial DNA in oxidatively deficient fibers of chronic obstructive pulmonary disease locomotor muscle', *Skeletal Muscle*, 6, p. 10.

Koonin, E.V. (2010) 'The origin and early evolution of eukaryotes in the light of phylogenomics', *Genome Biology*, 11(5).

Krajden, M., Minor, J.M., Rifkin, O. and Comanor, L. (1999) 'Effect of Multiple Freeze-Thaw Cycles on Hepatitis B Virus DNA and Hepatitis C Virus RNA Quantification as Measured with Branched-DNA Technology', *Journal of Clinical Microbiology*, 37(6), pp. 1683-1686.

Kraytsberg, Y., Kudryavtseva, E., McKee, A.C., Geula, C., Kowall, N.W. and Khrapko, K. (2006) 'Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons', *Nature Genetics*, 38(5), pp. 518-520.

Krishnan, K.J., Bender, A., Taylor, R.W. and Turnbull, D.M. (2007) 'A multiplex real-time PCR method to detect and quantify mitochondrial DNA deletions in individual cells', *Analytical Biochemistry*, 370(1), pp. 127-129.

Krishnan, K.J., Reeve, A.K., Samuels, D.C., Chinnery, P.F., Blackwood, J.K., Taylor, R.W., Wanrooij, S., Spelbrink, J.N., Lightowlers, R.N. and Turnbull, D.M. (2008) 'What causes mitochondrial DNA deletions in human cells?', *Nature Genetics*, 40(3), pp. 275-279.

Kühlbrandt, W. (2015) 'Structure and function of mitochondrial membrane protein complexes', *BMC Biology*, 13(1), p. 89.

Kulkarni, J., Gavrilidis, E., Worsley, R. and Hayes, E. (2012) 'Role of Estrogen Treatment in the Management of Schizophrenia', *CNS Drugs*, 26(7), pp. 549-557.

Lakatos, A., Derbeneva, O., Younes, D., Keator, D., Bakken, T., Lvova, M., Brandon, M., Guffanti, G., Reglodi, D., Saykin, A., Weiner, M., Macciardi, F., Schork, N., Wallace, D.C. and Potkin, S.G. (2010) 'Association between mitochondrial DNA variations and Alzheimer's disease in the ADNI cohort', *Neurobiology of Aging*, 31(8), pp. 1355-1363.

Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., Russo, G., Thorton-Wells, T.A., Jones, N., Smith, A.V., Chouraki, V., Thomas, C., Ikram, M.A., Zelenika, D., Vardarajan, B.N., Kamatani, Y., Lin, C.F., Gerrish, A., Schmidt, H., Kunkle, B., Dunstan, M.L., Ruiz, A., Bihoreau, M.T., Choi, S.H., Reitz, C., Pasquier, F., Cruchaga, C., Craig, D., Amin, N., Berr, C., Lopez, O.L., De Jager, P.L., Deramecourt, V., Johnston, J.A., Evans, D., Lovestone, S., Letenneur, L., Moron, F.J., Rubinsztein, D.C., Eiriksdottir, G., Sleegers, K., Goate, A.M., Fievet, N., Huentelman, M.W., Gill, M., Brown, K., Kamboh, M.I., Keller, L., Barberger-Gateau, P., McGuiness, B., Larson, E.B., Green, R., Myers, A.J., Dufouil, C., Todd, S., Wallon, D., Love, S., Rogaeva, E., Gallacher, J., St George-Hyslop, P., Clarimon, J., Lleo, A., Bayer, A., Tsuang, D.W., Yu, L., Tsolaki, M., Bossu, P., Spalletta, G., Proitsi, P., Collinge, J., Sorbi, S., Sanchez-Garcia, F., Fox, N.C., Hardy, J., Deniz Naranjo, M.C., Bosco, P., Clarke, R., Brayne, C., Galimberti, D., Mancuso, M., Matthews, F., European Alzheimer's Disease, I., Genetic, Environmental Risk in Alzheimer's, D., Alzheimer's Disease Genetic, C., Cohorts for, H., Aging Research in Genomic, E.,

Moebus, S., Mecocci, P., Del Zompo, M., Maier, W., Hampel, H., Pilotto, A., Bullido, M., Panza, F., Caffarra, P., et al. (2013) 'Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease', *Nat Genet*, 45(12), pp. 1452-8.

Lan, Q., Lim, U., Liu, C.S., Weinstein, S.J., Chanock, S., Bonner, M.R., Virtamo, J., Albanes, D. and Rothman, N. (2008) 'A prospective study of mitochondrial DNA copy number and risk of non-Hodgkin lymphoma', *Blood*, 112(10), pp. 4247-4249.

Lane, N. and Martin, W. (2010a) 'The energetics of genome complexity', *Nature*, 467(7318), pp. 929-34.

Lane, N. and Martin, W. (2010b) 'The energetics of genome complexity', *Nature*, 467(7318), pp. 929-934.

Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M. and Gray, M.W. (1997) 'An ancestral mitochondrial DNA resembling a eubacterial genome in miniature', *Nature*, 387(6632), pp. 493-497.

Lang, M., Sazzini, M., Calabrese, F.M., Simone, D., Boattini, A. and Romeo, G. (2012) 'Polymorphic NumtS trace human population relationships', *Hum Genet*, 131.

Larsson, N.-G., Wang, J., Wilhelmsson, H., Oldfors, A., Rustin, P., Lewandoski, M., Barsh, G.S. and Clayton, D.A. (1998) 'Mitochondrial transcription factor A is necessary for mtDNA maintance and embryogenesis in mice', *Nat Genet*, 18(3), pp. 231-236.

Laubenthal, L., Hoelker, M., Frahm, J., Dänicke, S., Gerlach, K., Südekum, K.H., Sauerwein, H. and Häussler, S. (2016) 'Mitochondrial DNA copy number and biogenesis in different tissues of early- and late-lactating dairy cows', *Journal of Dairy Science*, 99(2), pp. 1571-1583.

LeBlanc, D. (2004) *Statistics; Concepts and Applications for Science*. Sudbury, MA, USA: Jones and Bartlett Publishers.

Ledoux, S.P., Wilson, G.L., Beecham, E.J., Stevnsner, T., Wassermann, K. and Bohr, V.A. (1992) 'Repair of mitochondrial DNA after various types of DNA damage in chinese hamster ovary cells', *Carcinogenesis*, 13(11), pp. 1967-1973.

Lee, H.C. and Wei, Y.H. (2000) 'Mitochondrial role in life and death of the cell', *J Biomed Sci*, 7(1), pp. 2-15.

Lee, H.C., Yin, P.H., Lin, J.C., Wu, C.C., Chen, C.Y. and Wu, C.W. (2005) 'Mitochondrial genome instability and mtDNA depletion in human cancers', *Ann N Y Acad Sci*, 1042.

Lee, J.V., Carrer, A., Shah, S., Snyder, N.W., Wei, S., Venneti, S., Worth, A.J., Yuan, Z.F., Lim, H.W., Liu, S., Jackson, E., Aiello, N.M., Haas, N.B., Rebbeck, T.R., Judkins, A., Won, K.J., Chodosh, L.A., Garcia, B.A., Stanger, B.Z., Feldman, M.D., Blair, I.A. and Wellen, K.E. (2014a) 'Akt-dependent metabolic reprogramming regulates tumor cell Histone acetylation', *Cell Metabolism*, 20(2), pp. 306-319.

Lee, J.W., Park, K.D., Im, J.A., Kim, M.Y. and Lee, D.C. (2010) 'Mitochondrial DNA copy number in peripheral blood is associated with cognitive function in apparently healthy elderly women', *Clin Chim Acta*, 411(7-8), pp. 592-6.

Lee, J.Y., Lee, D.C., Im, J.A. and Lee, J.W. (2014b) 'Mitochondrial DNA copy number in peripheral blood is independently associated with visceral fat accumulation in healthy young adults', *International Journal of Endocrinology*, 2014.

Lee, W., Johnson, J., Gough, D.J., Donoghue, J., Cagnone, G.L.M., Vaghjiani, V., Brown, K.A., Johns, T.G. and St. John, J.C. (2015) 'Mitochondrial DNA copy number is regulated by DNA methylation and demethylation of POLGA in stem and cancer cells and their differentiated progeny', *Cell Death Dis*, 6, p. e1664.

Lefai, E., Fernandez-Moreno, M.A., Alahari, A., Kaguni, L.S. and Garesse, R. (2000) 'Differential regulation of the catalytic and accessory subunit genes of Drosophila mitochondrial DNA polymerase', *Journal of Biological Chemistry*, 275(42), pp. 33123-33133.

Lehman, J.J., Barger, P.M., Kovacs, A., Saffitz, J.E., Medeiros, D.M. and Kelly, D.P. (2000) 'Peroxisome proliferator–activated receptor γ coactivator-1 promotes cardiac mitochondrial biogenesis', *Journal of Clinical Investigation*, 106(7), pp. 847-856.

Lemmon, G.H. and Gardner, S.N. (2008) 'Predicting the sensitivity and specificity of published real-time PCR assays', *Annals of Clinical Microbiology and Antimicrobials*, 7, pp. 18-18.

Leonard, J.V. and Schapira, A.H.V. (2000) 'Mitochondrial respiratory chain disorders I: Mitochondrial DNA defects', *Lancet*, 355(9200), pp. 299-304.

Létourneau, D., Bédard, M., Cabana, J., Lefebvre, A., LeHoux, J.-G. and Lavigne, P. (2016) 'STARD6 on steroids: solution structure, multiple timescale backbone dynamics and ligand binding mechanism', *Scientific Reports*, 6, p. 28486.

Lewis, S.C., Uchiyama, L.F. and Nunnari, J. (2016) 'ER-mitochondria contacts couple mtDNA synthesis with mitochondrial division in human cells', *Science*, 353(6296).

Lezza, A.M.S. (2012) 'Mitochondrial transcription factor A (TFAM): one actor for different roles', *Frontiers in Biology*, 7(1), pp. 30-39.

Li, H., Handsaker, B. and Wysoker, A. (2009) 'The sequence alignment/Map format and SAMtools', *Bioinformatics*, 25.

Li, L., Hann, H.-W., Wan, S., Hann, R.S., Wang, C., Lai, Y., Ye, X., Evans, A., Myers, R.E., Ye, Z., Li, B., Xing, J. and Yang, H. (2016) 'Cell-free circulating mitochondrial DNA content and risk of hepatocellular carcinoma in patients with chronic HBV infection', 6, p. 23992.

Li, L.Y., Luo, X. and Wang, X. (2001) 'Endonuclease G is an apoptotic DNase when released from mitochondria', *Nature*, 412(6842), pp. 95-99.

Lim, E.L., Hollingsworth, K.G., Aribisala, B.S., Chen, M.J., Mathers, J.C. and Taylor, R. (2011) 'Reversal of type 2 diabetes: normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol', *Diabetologia*, 54(10), pp. 2506-2514.

Limongelli, A. and Tiranti, V. (2002) 'Inherited Mendelian defects of nuclear–mitochondrial communication affecting the stability of mitochondrial DNA', *Mitochondrion*, 2(1–2), pp. 39-46.

Lin, C.-S., Lee, H.-T., Lee, S.-Y., Shen, Y.-A., Wang, L.-S., Chen, Y.-J. and Wei, Y.-H. (2012) 'High Mitochondrial DNA Copy Number and Bioenergetic Function Are Associated with Tumor Invasion of Esophageal Squamous Cell Carcinoma Cell Lines', *International Journal of Molecular Sciences*, 13(9), pp. 11228-11246.

Liochev, S.I. (2013) 'Reactive oxygen species and the free radical theory of aging', *Free Radic Biol Med*, 60, pp. 1-4.

Liou, C.W., Chen, J.B., Tiao, M.M., Weng, S.W., Huang, T.L., Chuang, J.H., Chen, S.D., Chuang, Y.C., Lee, W.C., Lin, T.K. and Wang, P.W. (2012) 'Mitochondrial DNA

coding and control region variants as genetic risk factors for type 2 diabetes', *Diabetes*, 61(10), pp. 2642-51.

Liu, C.S., Tsai, C.S., Kuo, C.L., Chen, H.W., Lii, C.K., Ma, Y.S. and Wei, Y.H. (2003) 'Oxidative stress-related alteration of the copy number of mitochondrial DNA in human leukocytes', *Free Radical Research*, 37(12), pp. 1307-1317.

Liu, J., Cai, X., Xie, L., Tang, Y., Cheng, J., Wang, J., Wang, L. and Gong, J. (2015a) 'Circulating Cell Free Mitochondrial DNA is a Biomarker in the Development of Coronary Heart Disease in the Patients with Type 2 Diabetes', *Clin Lab*, 61(7), pp. 661-7.

Liu, J.Z., Tozzi, F., Waterworth, D.M., Pillai, S.G., Muglia, P., Middleton, L., Berrettini, W., Knouff, C.W., Yuan, X., Waeber, G., Vollenweider, P., Preisig, M., Wareham, N.J., Zhao, J.H., Loos, R.J.F., Barroso, I., Khaw, K.T., Grundy, S., Barter, P., Mahley, R., Kesaniemi, A., McPherson, R., Vincent, J.B., Strauss, J., Kennedy, J.L., Farmer, A., McGuffin, P., Day, R., Matthews, K., Bakke, P., Gulsvik, A., Lucae, S., Ising, M., Brueckl, T., Horstmann, S., Wichmann, H.E., Rawal, R., Dahmen, N., Lamina, C., Polasek, O., Zgaga, L., Huffman, J., Campbell, S., Kooner, J., Chambers, J.C., Burnett, M.S., Devaney, J.M., Pichard, A.D., Kent, K.M., Satler, L., Lindsay, J.M., Waksman, R., Epstein, S., Wilson, J.F., Wild, S.H., Campbell, H., Vitart, V., Reilly, M.P., Li, M., Qu, L., Wilensky, R., Matthai, W., Hakonarson, H.H., Rader, D.J., Franke, A., Wittig, M., Schäfer, A., Uda, M., Terracciano, A., Xiao, X., Busonero, F., Scheet, P., Schlessinger, D., Clair, D.S., Rujescu, D., Abecasis, G.R., Grabe, H.J., Teumer, A., Völzke, H., Petersmann, A., John, U., Rudan, I., Hayward, C., Wright, A.F., Kolcic, I., Wright, B.J., Thompson, J.R., Balmforth, A.J., Hall, A.S., Samani, N.J., Anderson, C.A., Ahmad, T., Mathew, C.G., Parkes, M., Satsangi, J., Caulfield, M., Munroe, P.B., Farrall, M., Dominiczak, A., Worthington, J., et al. (2010) 'Meta-analysis and imputation refines the association of 15q25 with smoking quantity', *Nature Genetics*, 42(5), pp. 436-440.

Liu, T.F., Vachharajani, V., Millet, P., Bharadwaj, M.S., Molina, A.J. and McCall, C.E. (2015b) 'Sequential Actions of SIRT1-RELB-SIRT3 Coordinate Nuclear-Mitochondrial Communication during Immunometabolic Adaptation to Acute Inflammation and Sepsis', *Journal of Biological Chemistry*, 290(1), pp. 396-408.

Livy, A., Lye, S., Jagdish, C.K., Hanis, N., Sharmila, V., Ler, L.W. and Pramod, B. (2012) 'Evaluation of Quality of DNA Extracted from Buccal Swabs for Microarray Based Genotyping', *Indian Journal of Clinical Biochemistry*, 27(1), pp. 28-33.

Lobatón, C.D., Vay, L., Hernández-SanMiguel, E., SantoDomingo, J., Moreno, A., Montero, M. and Alvarez, J. (2005) 'Modulation of mitochondrial Ca(2+) uptake by estrogen receptor agonists and antagonists', *British Journal of Pharmacology*, 145(7), pp. 862-871.

Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., Croteau-Chonka, D.C., Esko, T., Fall, T., Ferreira, T., Gustafsson, S., Kutalik, Z., Luan, J.a., Magi, R., Randall, J.C., Winkler, T.W., Wood, A.R., Workalemahu, T., Faul, J.D., Smith, J.A., Hua Zhao, J., Zhao, W., Chen, J., Fehrmann, R., Hedman, A.K., Karjalainen, J., Schmidt, E.M., Absher, D., Amin, N., Anderson, D., Beekman, M., Bolton, J.L., Bragg-Gresham, J.L., Buyske, S., Demirkan, A., Deng, G., Ehret, G.B., Feenstra, B., Feitosa, M.F., Fischer, K., Goel, A., Gong, J., Jackson, A.U., Kanoni, S., Kleber, M.E., Kristiansson, K., Lim, U., Lotay, V., Mangino, M., Mateo Leach, I., Medina-Gomez, C., Medland, S.E., Nalls, M.A., Palmer, C.D., Pasko, D., Pechlivanis, S., Peters, M.J., Prokopenko, I., Shungin, D., Stancakova, A., Strawbridge, R.J., Ju Sung, Y., Tanaka, T., Teumer, A., Trompet, S., van der Laan, S.W., van Setten, J., Van Vliet-Ostaptchouk, J.V., Wang, Z., Yengo, L., Zhang, W., Isaacs, A., Albrecht, E., Arnlov, J., Arscott, G.M., Attwood, A.P., Bandinelli, S., Barrett, A., Bas, I.N., Bellis, C., Bennett, A.J., Berne, C., Blagieva, R., Bluher, M., Bohringer, S., Bonnycastle, L.L., Bottcher, Y., Boyd, H.A., Bruinenberg, M., Caspersen, I.H., Ida Chen, Y.-D., Clarke, R., Warwick Daw, E., de Craen, A.J.M., Delgado, G., Dimitriou, M., et al. (2015) 'Genetic studies of body mass index yield new insights for obesity biology', *Nature*, 518(7538), pp. 197-206.

Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., Durbin, R. and L Price, A. (2016) 'Reference-based phasing using the Haplotype Reference Consortium panel', *Nat Genet*, 48(11), pp. 1443-1448.

Long, T., Hicks, M., Yu, H.-C., Biggs, W.H., Kirkness, E.F., Menni, C., Zierer, J., Small, K.S., Mangino, M., Messier, H., Brewerton, S., Turpaz, Y., Perkins, B.A., Evans, A.M., Miller, L.A.D., Guo, L., Caskey, C.T., Schork, N.J., Garner, C., Spector, T.D., Venter, J.C. and Telenti, A. (2017) 'Whole-genome sequencing identifies

common-to-rare variants associated with human blood metabolites', *Nat Genet*, 49(4), pp. 568-578.

López, S., Buil, A., Souto, J.C., Casademont, J., Blangero, J., Martinez-Perez, A., Fontcuberta, J., Lathrop, M., Almasy, L. and Soria, J.M. (2012) 'Sex-Specific Regulation of Mitochondrial DNA Levels: Genome-Wide Linkage Analysis to Identify Quantitative Trait Loci', *PLoS ONE*, 7(8), p. e42711.

Lorente, L., Iceta, R., Martín, M.M., López-Gallardo, E., Solé-Violán, J., Blanquer, J., Labarta, L., Díaz, C., Borreguero-León, J.M., Jiménez, A., Montoya, J. and Ruiz-Pesini, E. (2013) 'Severe Septic Patients with Mitochondrial DNA Haplogroup JT Show Higher Survival Rates: A Prospective, Multicenter, Observational Study', *PLOS ONE*, 8(9), p. e73320.

Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., Procaccio, V. and Wallace, D.C. (2013) 'mtDNA variation and analysis using MITOMAP and MITOMASTER', *Curr Protoc Bioinformatics*, 1.

Lu, B., Poirier, C., Gaspar, T., Gratzke, C., Harrison, W., Busija, D., Matzuk, M.M., Andersson, K.-E., Overbeek, P.A. and Bishop, C.E. (2008) 'A Mutation in the Inner Mitochondrial Membrane Peptidase 2-Like Gene (Immp2l) Affects Mitochondrial Function and Impairs Fertility in Mice1', *Biology of Reproduction*, 78(4), pp. 601-610.

Lynch, L.A., O'Connell, J.M., Kwasnik, A.K., Cawood, T.J., O'Farrelly, C. and O'Shea, D.B. (2009) 'Are natural killer cells protecting the metabolically healthy obese patient?', *Obesity*, 17.

Lynch, M., Koskella, B. and Schaack, S. (2006) 'Mutation pressure and the evolution of organelle genomic architecture', *Science*, 311(5768), pp. 1727-1730.

Ma, L. and Chung, W.K. (2014) 'Quantitative Analysis of Copy Number Variants Based on Real-Time LightCycler PCR', *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]*, 80, pp. 7.21.1-7.21.8.

Ma, Y.Y., Wu, T.F., Liu, Y.P., Wang, Q., Li, X.Y., Song, J.Q., Shi, X.Y., Zhang, W.N., Zhao, M., Hu, L.Y., Yang, Y.L. and Zou, L.P. (2013) 'Heterogeneity of six children and their mothers with mitochondrial DNA 3243 A>G mutation', *Mitochondrial DNA*, 24(3), pp. 297-302.

Maagaard, A., Holberg-Petersen, M., Kollberg, G., Oldfors, A., Sandvik, L. and Bruun, J.N. (2006) 'Mitochondrial (mt)DNA changes in tissue may not be reflected by depletion of mtDNA in peripheral blood mononuclear cells in HIV-infected patients', *Antivir Ther*, 11(5), pp. 601-8.

Maassen, J.A., ‘t Hart, L.M., van Essen, E., Heine, R.J., Nijpels, G., Jahangir Tafrechi, R.S., Raap, A.K., Janssen, G.M.C. and Lemkes, H.H.P.J. (2004) 'Mitochondrial Diabetes', *Molecular Mechanisms and Clinical Presentation*, 53(suppl 1), pp. S103-S109.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., MayPendlington, Z., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F. and Parkinson, H. (2017) 'The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog)', *Nucleic Acids Research*, 45(D1), pp. D896-D901.

Magner, M., Szentiványi, K., Švandová, I., Ješina, P., Tesařová, M., Honzík, T. and Zeman, J. (2011) 'Elevated CSF-lactate is a reliable marker of mitochondrial disorders in children even after brief seizures', *European Journal of Paediatric Neurology*, 15(2), pp. 101-108.

Malik, A.N. and Czajka, A. (2013) 'Is mitochondrial DNA content a potential biomarker of mitochondrial dysfunction?', *Mitochondrion*, 13(5), pp. 481-492.

Malik, A.N., Czajka, A. and Cunningham, P. (2016) 'Accurate quantification of mouse mitochondrial DNA without co-amplification of nuclear mitochondrial insertion sequences', *Mitochondrion*, 29, pp. 59-64.

Malik, A.N., Shahni, R. and Iqbal, M.M. (2009) 'Increased peripheral blood mitochondrial DNA in type 2 diabetic patients with nephropathy', *Diabetes Research and Clinical Practice*, 86(2), pp. e22-e24.

Malik, A.N., Shahni, R., Rodriguez-de-Ledesma, A., Laftah, A. and Cunningham, P. (2011) 'Mitochondrial DNA as a non-invasive biomarker: Accurate quantification using real time quantitative PCR without co-amplification of pseudogenes and dilution bias', *Biochemical and Biophysical Research Communications*, 412(1), pp. 1-7.

Mambo, E., Chatterjee, A., Xing, M., Tallini, G., Haugen, B.R., Yeung, S.C., Sukumar, S. and Sidransky, D. (2005) 'Tumor-specific changes in mtDNA content in human cancer', *Int J Cancer*, 116.

Man, P.Y.W., Griffiths, P.G., Brown, D.T., Howell, N., Turnbull, D.M. and Chinnery, P.F. (2003) 'The epidemiology of leber hereditary optic neuropathy in the North East of England', *American Journal of Human Genetics*, 72(2), pp. 333-339.

Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T., Bielak, L.F., Prokopenko, I., Amin, N., Barnes, D., Cadby, G., Hottenga, J.J., Ingelsson, E., Jackson, A.U., Johnson, T., Kanoni, S., Ladenvall, C., Lagou, V., Lahti, J., Lecoeur, C., Liu, Y., Martinez-Larrad, M.T., Montasser, M.E., Navarro, P., Perry, J.R., Rasmussen-Torvik, L.J., Salo, P., Sattar, N., Shungin, D., Strawbridge, R.J., Tanaka, T., van Duijn, C.M., An, P., de Andrade, M., Andrews, J.S., Aspelund, T., Atalay, M., Aulchenko, Y., Balkau, B., Bandinelli, S., Beckmann, J.S., Beilby, J.P., Bellis, C., Bergman, R.N., Blangero, J., Boban, M., Boehnke, M., Boerwinkle, E., Bonnycastle, L.L., Boomsma, D.I., Borecki, I.B., Bottcher, Y., Bouchard, C., Brunner, E., Budimir, D., Campbell, H., Carlson, O., Chines, P.S., Clarke, R., Collins, F.S., Corbaton-Anchuelo, A., Couper, D., de Faire, U., Dedoussis, G.V., Deloukas, P., Dimitriou, M., Egan, J.M., Eiriksdottir, G., Erdos, M.R., Eriksson, J.G., Eury, E., Ferrucci, L., Ford, I., Forouhi, N.G., Fox, C.S., Franzosi, M.G., Franks, P.W., Frayling, T.M., Froguel, P., Galan, P., de Geus, E., Gigante, B., Glazer, N.L., Goel, A., Groop, L., Gudnason, V., Hallmans, G., Hamsten, A., Hansson, O., Harris, T.B., Hayward, C., Heath, S., Hercberg, S., Hicks, A.A., Hingorani, A., Hofman, A., Hui, J., Hung, J., et al. (2012) 'A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance', *Nat Genet*, 44(6), pp. 659-69.

Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) 'A new multipoint method for genome-wide association studies by imputation of genotypes', *Nat Genet*, 39(7), pp. 906-13.

Margulis, L. (1971) 'Symbiosis and evolution', *Scientific American*, 225(2), pp. 48-57.

Marouli, E., Graff, M., Medina-Gomez, C., Lo, K.S., Wood, A.R., Kjaer, T.R., Fine, R.S., Lu, Y., Schurmann, C., Highland, H.M., Rüeger, S., Thorleifsson, G., Justice, A.E., Lamparter, D., Stirrups, K.E., Turcot, V., Young, K.L., Winkler, T.W., Esko, T.,

Karaderi, T., Locke, A.E., Masca, N.G.D., Ng, M.C.Y., Mudgal, P., Rivas, M.A., Vedantam, S., Mahajan, A., Guo, X., Abecasis, G., Aben, K.K., Adair, L.S., Alam, D.S., Albrecht, E., Allin, K.H., Allison, M., Amouyel, P., Appel, E.V., Arveiler, D., Asselbergs, F.W., Auer, P.L., Balkau, B., Banas, B., Bang, L.E., Benn, M., Bergmann, S., Bielak, L.F., Blüher, M., Boeing, H., Boerwinkle, E., Böger, C.A., Bonnycastle, L.L., Bork-Jensen, J., Bots, M.L., Bottinger, E.P., Bowden, D.W., Brandslund, I., Breen, G., Brilliant, M.H., Broer, L., Burt, A.A., Butterworth, A.S., Carey, D.J., Caulfield, M.J., Chambers, J.C., Chasman, D.I., Chen, Y.-D.I., Chowdhury, R., Christensen, C., Chu, A.Y., Cocca, M., Collins, F.S., Cook, J.P., Corley, J., Galbany, J.C., Cox, A.J., Cuellar-Partida, G., Danesh, J., Davies, G., de Bakker, P.I.W., de Borst, G.J., de Denus, S., de Groot, M.C.H., de Mutsert, R., Deary, I.J., Dedoussis, G., Demerath, E.W., den Hollander, A.I., Dennis, J.G., Di Angelantonio, E., Drenos, F., Du, M., Dunning, A.M., Easton, D.F., Ebeling, T., Edwards, T.L., Ellinor, P.T., Elliott, P., Evangelou, E., Farmaki, A.-E., Faul, J.D., et al. (2017) 'Rare and low-frequency coding variants alter human adult height', *Nature*, 542(7640), pp. 186-190.

Martin, P., McGovern, A., Orozco, G., Duffus, K., Yarwood, A., Schoenfelder, S., Cooper, N.J., Barton, A., Wallace, C., Fraser, P., Worthington, J. and Eyre, S. (2015) 'Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci', *Nat Commun*, 6, p. 10069.

Mattingly, K.A., Ivanova, M.M., Riggs, K.A., Wickramasinghe, N.S., Barch, M.J. and Klinge, C.M. (2008) 'Estradiol Stimulates Transcription of Nuclear Respiratory Factor-1 and Increases Mitochondrial Biogenesis', *Molecular Endocrinology*, 22(3), pp. 609-622.

Mauvais-Jarvis, F., Arnold, A.P. and Reue, K. (2017) 'A Guide for the Design of Pre-clinical Studies on Sex Differences in Metabolism', *Cell Metabolism*, 25(6), pp. 1216-1230.

Mayo Clinic Staff (2016) *Complete blood count (CBC)*. Available at: http://www.mayoclinic.org/tests-procedures/complete-blood-count/details/results/rsc-20257186 (Accessed: 4th August).

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P.A. and Hirschhorn, J.N. (2008) 'Genome-wide association studies for complex traits: consensus, uncertainty and challenges', *Nat Rev Genet*, 9(5), pp. 356-369.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L.J., Zhang, H., Mahajan, A., Veldink, J., Peters, U., Pato, C., van Duijn, C.M., Gillies, C.E., Gandin, I., Mezzavilla, M., Gilly, A., Cocca, M., Traglia, M., Angius, A., Barrett, J.C., Boomsma, D., Branham, K., Breen, G., Brummett, C.M., Busonero, F., Campbell, H., Chan, A., Chen, S., Chew, E., Collins, F.S., Corbin, L.J., Smith, G.D., Dedoussis, G., Dorr, M., Farmaki, A.E., Ferrucci, L., Forer, L., Fraser, R.M., Gabriel, S., Levy, S., Groop, L., Harrison, T., Hattersley, A., Holmen, O.L., Hveem, K., Kretzler, M., Lee, J.C., McGue, M., Meitinger, T., Melzer, D., Min, J.L., Mohlke, K.L., Vincent, J.B., Nauck, M., Nickerson, D., Palotie, A., Pato, M., Pirastu, N., McInnis, M., Richards, J.B., Sala, C., Salomaa, V., Schlessinger, D., Schoenherr, S., Slagboom, P.E., Small, K., Spector, T., Stambolian, D., Tuke, M., Tuomilehto, J., Van den Berg, L.H., Van Rheenen, W., Volker, U., Wijmenga, C., Toniolo, D., Zeggini, E., Gasparini, P., Sampson, M.G., Wilson, J.F., Frayling, T., de Bakker, P.I., Swertz, M.A., McCarroll, S., Kooperberg, C., Dekker, A., Altshuler, D., Willer, C., Iacono, W., Ripatti, S., et al. (2016) 'A reference panel of 64,976 haplotypes for genotype imputation', *Nat Genet*, 48(10), pp. 1279-83.

McCarthy, S., Das, S., Kretzschmar, W., Durbin, R., Abecasis, G. and Marchini, J. (2015) 'A reference panel of 64,976 haplotypes for genotype imputation', *bioRxiv*.

McCutcheon, J.P., McDonald, B.R. and Moran, N.A. (2009) 'Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont', *PLoS Genetics*, 5(7).

McDermott-Roe, C., Ye, J., Ahmed, R., Sun, X.M., Serafín, A., Ware, J., Bottolo, L., Muckett, P., Cañas, X., Zhang, J., Rowe, G.C., Buchan, R., Lu, H., Braithwaite, A., Mancini, M., Hauton, D., Martí, R., Garcí-Arumí, E., Hubner, N., Jacob, H., Serikawa, T., Zidek, V., Papousek, F., Kolar, F., Cardona, M., Ruiz-Meana, M., García-Dorado, D., Comella, J.X., Felkin, L.E., Barton, P.J.R., Arany, Z., Pravenec, M., Petretto, E., Sanchis, D. and Cook, S.A. (2011) 'Endonuclease G is a novel determinant of cardiac hypertrophy and mitochondrial function', *Nature*, 478(7367), pp. 114-118.

McKenzie, M., Trounce, I.A., Cassar, C.A. and Pinkert, C.A. (2004) 'Production of homoplasmic xenomitochondrial mice', *Proc Natl Acad Sci U S A*, 101(6), pp. 1685-90.

Mengel-From, J., Thinggaard, M., Dalgård, C., Kyvik, K.O., Christensen, K. and Christiansen, L. (2014) 'Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly', *Human Genetics*, 133(9), pp. 1149-1159.

Meyer, J.N. and Bess, A.S. (2012) 'Involvement of autophagy and mitochondrial dynamics in determining the fate and effects of irreparable mitochondrial DNA damage', *Autophagy*, 8(12), pp. 1822-1823.

Meyer, J.N., Leung, M.C.K., Rooney, J.P., Sendoel, A., Hengartner, M.O., Kisby, G.E. and Bess, A.S. (2013) 'Mitochondria as a Target of Environmental Toxicants', *Toxicological Sciences*, 134(1), pp. 1-17.

Mi, J., Tian, G., Liu, S., Li, X., Ni, T., Zhang, L. and Wang, B. (2015) 'The relationship between altered mitochondrial DNA copy number and cancer risk: A meta-analysis', *Scientific Reports*, 5.

Michaels, G.S., Hauswirth, W.W. and Laipis, P.J. (1982) 'Mitochondrial DNA copy number in bovine oocytes and somatic cells', *Developmental Biology*, 94(1), pp. 246-251.

Milenkovic, D., Matic, S., kühl, I., Ruzzenente, B., Freyer, C., Jemt, E., Park, C.B., Falkenberg, M. and Larsson, N. (2013) 'Twinkle is an essential mitochondrial helicase required for synthesis of nascent D-loop strands and complete mtDNA replication', *Human Molecular Genetics*, 22(10), pp. 1983-1993.

Miller, F.J., Rosenfeldt, F.L., Zhang, C., Linnane, A.W. and Nagley, P. (2003) 'Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age', *Nucleic Acids Res*, 31.

Miller, W.L. 20 (2010) 'Role of mitochondria in steroidogenesis' *Endocrine Development* [Article]. pp. 1-19. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-

78751550536&doi=10.1159%2f000321204&partnerID=40&md5=3e03cb2fd97b6ef72cb4e15cfee52990.

Miller, W.L. (2013) 'Steroid hormone synthesis in mitochondria', *Molecular and Cellular Endocrinology*, 379(1), pp. 62-73.

Miller, W.L. and Auchus, R.J. (2011) 'The Molecular Biology, Biochemistry, and Physiology of Human Steroidogenesis and Its Disorders', *Endocrine Reviews*, 32(1), pp. 81-151.

Mills, E.L., Kelly, B. and O'Neill, L.A.J. (2017) 'Mitochondria are the powerhouses of immunity', *Nat Immunol*, 18(5), pp. 488-498.

Misawa, T., Takahama, M. and Saitoh, T. (2017) 'Mitochondria-Endoplasmic Reticulum Contact Sites Mediate Innate Immune Responses', *Adv Exp Med Biol*, 997, pp. 187-197.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easleyf, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A. and Wallace, D.C. (2003) 'Natural selection shaped regional mtDNA variation in humans', *Proceedings of the National Academy of Sciences of the United States of America*, 100(1), pp. 171-176.

Mishra, P., Carelli, V., Manfredi, G. and Chan, D.C. (2014) 'Proteolytic cleavage of Opa1 stimulates mitochondrial inner membrane fusion and couples fusion to oxidative phosphorylation', *Cell Metabolism*, 19(4), pp. 630-641.

Mishra, P. and Chan, D.C. (2014) 'Mitochondrial dynamics and inheritance during cell division, development and disease', *Nat Rev Mol Cell Biol*, 15(10), pp. 634-646.

Mitchell, A.L., Elson, J.L., Howell, N., Taylor, R.W. and Turnbull, D.M. (2006) 'Sequence variation in mitochondrial complex I genes: mutation or polymorphism?', *Journal of Medical Genetics*, 43(2), pp. 175-179.

Mitchell, P. (1961) 'Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism', *Nature*, 191(4784), pp. 144-148.

Miyazaki, T., Mori, S., Shigemoto, K., Larsson, N.-G., Nakamura, T., Kato, S., Nakashima, T., Takayanagi, H. and Tanaka, S. (2012) 'Maintenance of mitochondrial

DNA copy number is essential for osteoclast survival', *Arthritis Research & Therapy*, 14(1), p. P49.

Montoya, J., Christianson, T., Levens, D., Rabinowitz, M. and Attardi, G. (1982) 'Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 79(23 I), pp. 7195-7199.

Morel, Y. and Barouki, R. (1999) 'Repression of gene expression by oxidative stress', *Biochem J*, 342 Pt 3, pp. 481-96.

Morita, M., Gravel, S.-P., Chénard, V., Sikström, K., Zheng, L., Alain, T., Gandin, V., Avizonis, D., Arguello, M., Zakaria, C., McLaughlan, S., Nouet, Y., Pause, A., Pollak, M., Gottlieb, E., Larsson, O., St-Pierre, J., Topisirovic, I. and Sonenberg, N. (2013) 'mTORC1 Controls Mitochondrial Activity and Biogenesis through 4E-BP-Dependent Translational Regulation', *Cell Metabolism*, 18(5), pp. 698-711.

Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segre, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A., Prokopenko, I., Kang, H.M., Dina, C., Esko, T., Fraser, R.M., Kanoni, S., Kumar, A., Lagou, V., Langenberg, C., Luan, J., Lindgren, C.M., Muller-Nurasyid, M., Pechlivanis, S., Rayner, N.W., Scott, L.J., Wiltshire, S., Yengo, L., Kinnunen, L., Rossin, E.J., Raychaudhuri, S., Johnson, A.D., Dimas, A.S., Loos, R.J., Vedantam, S., Chen, H., Florez, J.C., Fox, C., Liu, C.T., Rybin, D., Couper, D.J., Kao, W.H., Li, M., Cornelis, M.C., Kraft, P., Sun, Q., van Dam, R.M., Stringham, H.M., Chines, P.S., Fischer, K., Fontanillas, P., Holmen, O.L., Hunt, S.E., Jackson, A.U., Kong, A., Lawrence, R., Meyer, J., Perry, J.R., Platou, C.G., Potter, S., Rehnberg, E., Robertson, N., Sivapalaratnam, S., Stancakova, A., Stirrups, K., Thorleifsson, G., Tikkanen, E., Wood, A.R., Almgren, P., Atalay, M., Benediktsson, R., Bonnycastle, L.L., Burtt, N., Carey, J., Charpentier, G., Crenshaw, A.T., Doney, A.S., Dorkhan, M., Edkins, S., Emilsson, V., Eury, E., Forsen, T., Gertow, K., Gigante, B., Grant, G.B., Groves, C.J., Guiducci, C., Herder, C., Hreidarsson, A.B., Hui, J., James, A., Jonsson, A., Rathmann, W., Klopp, N., Kravic, J., Krjutskov, K., Langford, C., Leander, K., Lindholm, E., Lobbens, S., Mannisto, S., et al. (2012) 'Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes', *Nat Genet*, 44(9), pp. 981-90.

Morrison, J. and Gauderman, W.J. (2002) *QUANTO* (Version 1.2.4) [Computer program]. University of Southern California. Available at: http://biostats.usc.edu/Quanto.html.

Mourier, T., Hansen, A.J., Willerslev, E. and Arctander, P. (2001) 'The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus', *Mol Biol Evol*, 18.

Mueller, E.E., Brunner, S.M., Mayr, J.A., Stanger, O., Sperl, W. and Kofler, B. (2012) 'Functional Differences between Mitochondrial Haplogroup T and Haplogroup H in HEK293 Cybrid Cells', *PLoS ONE*, 7(12), p. e52367.

Mueller, K.F., Meerpohl, J.J., Briel, M., Antes, G., Elm, E.v., Lang, B., Gloy, V., Motschall, E., Schwarzer, G. and Bassler, D. (2013) 'Detecting, quantifying and adjusting for publication bias in meta-analyses: protocol of a systematic review on methods', *Systematic Reviews*, 2, pp. 60-60.

Müller-Höcker, J., Horvath, R., Schäfer, S., Hessel, H., Müller-Felber, W., Kühr, J., Copeland, W.C. and Seibel, P. (2011) 'Mitochondrial DNA depletion and fatal infantile hepatic failure due to mutations in the mitochondrial polymerase γ (POLG) gene: A combined morphological/enzyme histochemical and immunocytochemical/biochemical and molecular genetic study', *Journal of Cellular and Molecular Medicine*, 15(2), pp. 445-456.

Muller, H.J. (1964) 'The Relation of Recombination to Mutational Advance', *Mutat Res*, 106, pp. 2-9.

Müller, M., Mentel, M., Van Hellemond, J.J., Henze, K., Woehle, C., Gould, S.B., Yu, R.Y., Van Der Giezen, M., Tielens, A.G.M. and Martin, W.F. (2012) 'Biochemistry and evolution of anaerobic energy metabolism in eukaryotes', *Microbiology and Molecular Biology Reviews*, 76(2), pp. 444-495.

Nagao, Y., Totsuka, Y., Atomi, Y., Kaneda, H., Lindahl, K.F., Imai, H. and Yonekawa, H. (1998) 'Decreased physical performance of congenic mice with mismatch between the nuclear and the mitochondrial genome', *Genes Genet Syst*, 73(1), pp. 21-7.

Nalls, M.A., Couper, D.J., Tanaka, T., van Rooij, F.J.A., Chen, M.-H., Smith, A.V., Toniolo, D., Zakai, N.A., Yang, Q., Greinacher, A., Wood, A.R., Garcia, M., Gasparini, P., Liu, Y., Lumley, T., Folsom, A.R., Reiner, A.P., Gieger, C., Lagou, V., Felix, J.F.,

Völzke, H., Gouskova, N.A., Biffi, A., Döring, A., Völker, U., Chong, S., Wiggins, K.L., Rendon, A., Dehghan, A., Moore, M., Taylor, K., Wilson, J.G., Lettre, G., Hofman, A., Bis, J.C., Pirastu, N., Fox, C.S., Meisinger, C., Sambrook, J., Arepalli, S., Nauck, M., Prokisch, H., Stephens, J., Glazer, N.L., Cupples, L.A., Okada, Y., Takahashi, A., Kamatani, Y., Matsuda, K., Tsunoda, T., Tanaka, T., Kubo, M., Nakamura, Y., Yamamoto, K., Kamatani, N., Stumvoll, M., Tönjes, A., Prokopenko, I., Illig, T., Patel, K.V., Garner, S.F., Kuhnel, B., Mangino, M., Oostra, B.A., Thein, S.L., Coresh, J., Wichmann, H.E., Menzel, S., Lin, J., Pistis, G., Uitterlinden, A.G., Spector, T.D., Teumer, A., Eiriksdottir, G., Gudnason, V., Bandinelli, S., Frayling, T.M., Chakravarti, A., van Duijn, C.M., Melzer, D., Ouwehand, W.H., Levy, D., Boerwinkle, E., Singleton, A.B., Hernandez, D.G., Longo, D.L., Soranzo, N., Witteman, J.C.M., Psaty, B.M., Ferrucci, L., Harris, T.B., O'Donnell, C.J. and Ganesh, S.K. (2011) 'Multiple Loci Are Associated with White Blood Cell Phenotypes', *PLoS Genetics*, 7(6), p. e1002113.

Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., Schulte, C., Keller, M.F., Arepalli, S., Letson, C., Edsall, C., Stefansson, H., Liu, X., Pliner, H., Lee, J.H., Cheng, R., International Parkinson's Disease Genomics, C., Parkinson's Study Group Parkinson's Research: The Organized, G.I., andMe, Gene, P.D., NeuroGenetics Research, C., Hussman Institute of Human, G., The Ashkenazi Jewish Dataset, I., Cohorts for, H., Aging Research in Genetic, E., North American Brain Expression, C., United Kingdom Brain Expression, C., Greek Parkinson's Disease, C., Alzheimer Genetic Analysis, G., Ikram, M.A., Ioannidis, J.P.A., Hadjigeorgiou, G.M., Bis, J.C., Martinez, M., Perlmutter, J.S., Goate, A., Marder, K., Fiske, B., Sutherland, M., Xiromerisiou, G., Myers, R.H., Clark, L.N., Stefansson, K., Hardy, J.A., Heutink, P., Chen, H., Wood, N.W., Houlden, H., Payami, H., Brice, A., Scott, W.K., Gasser, T., Bertram, L., Eriksson, N., Foroud, T. and Singleton, A.B. (2014) 'Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease', *Nature genetics*, 46(9), pp. 989-993.

Narendra, D.P., Jin, S.M., Tanaka, A., Suen, D.F., Gautier, C.A., Shen, J., Cookson, M.R. and Youle, R.J. (2010) 'PINK1 is selectively stabilized on impaired mitochondria to activate Parkin', *PLoS Biology*, 8(1).

Nargund, Amrita M., Fiorese, Christopher J., Pellegrino, Mark W., Deng, P. and Haynes, Cole M. (2015) 'Mitochondrial and Nuclear Accumulation of the Transcription

Factor ATFS-1 Promotes OXPHOS Recovery during the UPRmt', *Molecular Cell*, 58(1), pp. 123-133.

Nass, M.M. (1966) 'The circularity of mitochondrial DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 56(4), pp. 1215-1222.

Nass, M.M. and Nass, S. (1963a) 'INTRAMITOCHONDRIAL FIBERS WITH DNA CHARACTERISTICS. I. FIXATION AND', *The Journal of cell biology*, 19, pp. 593-611.

Nass, M.M.K. (1969) 'Mitochondrial DNA. I. Intramitochondrial distribution and structural relations of single- and double-length circular DNA', *Journal of Molecular Biology*, 42(3), pp. 521-"526,IN3-IN9,527-528".

Nass, M.M.K. (1973) 'Differential methylation of mitochondrial and nuclear DNA in cultured mouse, hamster and virus-transformed hamster cells In vivo and in vitro methylation', *Journal of Molecular Biology*, 80(1), pp. 155-175.

Nass, M.M.K. and Nass, S. (1962) 'Fibrous structures within the matrix of developing chick embryo mitochondria', *Experimental Cell Research*, 26(2), pp. 424-427.

Nass, M.M.K., Nass, S. and Afzelius, B.A. (1965) 'The general occurrence of mitochondrial DNA', *Experimental Cell Research*, 37(3), pp. 516-539.

Nass, S. and Nass, M.M.K. (1963b) 'AN ELECTRON HISTOCHEMICAL STUDY OF MITOCHONDRIAL FIBROUS INCLUSIONS', *Journal of the Royal Microscopical Society*, 81(3-4), pp. 209-213.

National Human Genome Research Institute (2003) *The Human Genome Project Completion: Frequently Asked Questions*. Available at: http://www.genome.gov/11006943 (Accessed: April 21).

Nemoto, S., Fergusson, M.M. and Finkel, T. (2005) 'SIRT1 Functionally Interacts with the Metabolic Regulator and Transcriptional Coactivator PGC-1α', *Journal of Biological Chemistry*, 280(16), pp. 16456-16460.

Nesbitt, V., Pitceathly, R.D.S., Turnbull, D.M., Taylor, R.W., Sweeney, M.G., Mudanohwo, E.E., Rahman, S., Hanna, M.G. and McFarland, R. (2013) 'The UK MRC Mitochondrial Disease Patient Cohort Study: clinical phenotypes associated

with the m.3243A>G mutation—implications for diagnosis and management', *Journal of Neurology, Neurosurgery & Psychiatry*, 84(8), pp. 936-938.

Ng, W.L. and Bassler, B.L. 43 (2009) 'Bacterial quorum-sensing network architectures' *Annual Review of Genetics*. pp. 197-222.

Ngo, H.B., Kaiser, J.T. and Chan, D.C. (2011) 'The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA', *Nat Struct Mol Biol*, 18(11), pp. 1290-1296.

NHS (2017) *NHS Choices*. Available at: http://www.nhs.uk/.

NHS Blood and Transplant. (2017) *Give Blood - Who Can Give Blood*. Available at: https://www.blood.co.uk/who-can-give-blood/ (Accessed: 6th August).

Nicklas, J.A., Brooks, E.M., Hunter, T.C., Single, R. and Branda, R.F. (2004) 'Development of a quantitative PCR (TaqMan) assay for relative mitochondrial DNA copy number and the common mitochondrial DNA deletion in the rat', *Environmental and Molecular Mutagenesis*, 44(4), pp. 313-320.

Niemi, A.-K., Hervonen, A., Hurme, M., Karhunen, P.J., Jylhä, M. and Majamaa, K. (2003) 'Mitochondrial DNA polymorphisms associated with longevity in a Finnish population', *Human Genetics*, 112(1), pp. 29-33.

Nikkanen, J., Forsström, S., Euro, L., Paetau, I., Kohnz, R.A., Wang, L., Chilov, D., Viinamäki, J., Roivainen, A., Marjamäki, P., Liljenbäck, H., Ahola, S., Buzkova, J., Terzioglu, M., Khan, N.A., Pirnes-Karhu, S., Paetau, A., Lönnqvist, T., Sajantila, A., Isohanni, P., Tyynismaa, H., Nomura, D.K., Battersby, B.J., Velagapudi, V., Carroll, C.J. and Suomalainen, A. (2016) 'Mitochondrial DNA Replication Defects Disturb Cellular dNTP Pools and Remodel One-Carbon Metabolism', *Cell Metabolism*, 23(4), pp. 635-648.

Nilsen, J., Chen, S., Irwin, R.W., Iwamoto, S. and Brinton, R.D. (2006) 'Estrogen protects neuronal cells from amyloid beta-induced apoptosis via regulation of mitochondrial proteins and function', *BMC Neuroscience*, 7, pp. 74-74.

Ning, Y., Bai, Q., Lu, H., Li, X., Pandak, W.M., Zhao, F., Chen, S., Ren, S. and Yin, L. (2009) 'Overexpression of mitochondrial cholesterol delivery protein, StAR,

decreases intracellular lipids and inflammatory factors secretion in macrophages', *Atherosclerosis*, 204(1), pp. 114-120.

Niranjan, B.G., Bhat, N.K. and Avadhani, N.G. (1982) 'Preferential attack of mitochondrial DNA by aflatoxin B1 during hepatocarcinogenesis', *Science*, 215(4528), pp. 73-75.

Nussbaum, R.L. and Ellis, C.E. (2003) 'Alzheimer's Disease and Parkinson's Disease', *New England Journal of Medicine*, 348(14), pp. 1356-1364.

O'Brien, T.W. (2002) 'Evolution of a protein-rich mitochondrial ribosome: Implications for human genetic disease', *Gene*, 286(1), pp. 73-79.

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J.E., Rudan, I., McQuillan, R., Fraser, R.M., Campbell, H., Polasek, O., Asiki, G., Ekoru, K., Hayward, C., Wright, A.F., Vitart, V., Navarro, P., Zagury, J.F., Wilson, J.F., Toniolo, D., Gasparini, P., Soranzo, N., Sandhu, M.S. and Marchini, J. (2014) 'A general approach for haplotype phasing across the full spectrum of relatedness', *PLoS Genet*, 10(4), p. e1004234.

O'Sullivan, S.T. and O'Connor, T.P.F. (1997) 'Immunosuppression following thermal injury: the pathogenesis of immunodysfunction', *British Journal of Plastic Surgery*, 50(8), pp. 615-623.

Okada, Y., Kamatani, Y., Takahashi, A., Matsuda, K., Hosono, N., Ohmiya, H., Daigo, Y., Yamamoto, K., Kubo, M., Nakamura, Y. and Kamatani, N. (2010) 'Common variations in PSMD3-CSF3 and PLCB4 are associated with neutrophil count', *Hum Mol Genet*, 19(10), pp. 2079-85.

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., Graham, R.R., Manoharan, A., Ortmann, W., Bhangale, T., Denny, J.C., Carroll, R.J., Eyler, A.E., Greenberg, J.D., Kremer, J.M., Pappas, D.A., Jiang, L., Yin, J., Ye, L., Su, D.F., Yang, J., Xie, G., Keystone, E., Westra, H.J., Esko, T., Metspalu, A., Zhou, X., Gupta, N., Mirel, D., Stahl, E.A., Diogo, D., Cui, J., Liao, K., Guo, M.H., Myouzen, K., Kawaguchi, T., Coenen, M.J., van Riel, P.L., van de Laar, M.A., Guchelaar, H.J., Huizinga, T.W., Dieude, P., Mariette, X., Bridges, S.L., Jr., Zhernakova, A., Toes, R.E., Tak, P.P., Miceli-Richard, C., Bang, S.Y., Lee, H.S., Martin, J., Gonzalez-Gay, M.A., Rodriguez-Rodriguez, L., Rantapaa-Dahlqvist, S.,

Arlestig, L., Choi, H.K., Kamatani, Y., Galan, P., Lathrop, M., consortium, R., consortium, G., Eyre, S., Bowes, J., Barton, A., de Vries, N., Moreland, L.W., Criswell, L.A., Karlson, E.W., Taniguchi, A., Yamada, R., Kubo, M., Liu, J.S., Bae, S.C., Worthington, J., Padyukov, L., Klareskog, L., Gregersen, P.K., Raychaudhuri, S., Stranger, B.E., De Jager, P.L., Franke, L., Visscher, P.M., Brown, M.A., Yamanaka, H., Mimori, T., Takahashi, A., Xu, H., Behrens, T.W., Siminovitch, K.A., Momohara, S., Matsuda, F., Yamamoto, K. and Plenge, R.M. (2014) 'Genetics of rheumatoid arthritis contributes to biology and drug discovery', *Nature*, 506(7488), pp. 376-81.

Olson, M.W., Wang, Y., Elder, R.H. and Kaguni, L.S. (1995) 'Subunit structure of mitochondrial DNA polymerase from Drosophila embryos: Physical and immunological studies', *Journal of Biological Chemistry*, 270(48), pp. 28932-28937.

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H. and O'Shea, C.C. (2017) 'ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells', *Science*, 357(6349).

Padovan-Merhar, O., Nair, G.P., Biaesch, A.G., Mayer, A., Scarfone, S., Foley, S.W., Wu, A.R., Churchman, L.S., Singh, A. and Raj, A. (2014) 'Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms', *Molecular Cell*.

Paech, K., Webb, P., Kuiper, G.G.J.M., Nilsson, S., Gustafsson, J.-Å., Kushner, P.J. and Scanlan, T.S. (1997) 'Differential Ligand Activation of Estrogen Receptors ERα and ERβ at AP1 Sites', *Science*, 277(5331), pp. 1508-1510.

Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S.A., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K., Hill, D.E., Vidal, M., Evans, J.G., Thorburn, D.R., Carr, S.A. and Mootha, V.K. (2008) 'A Mitochondrial Protein Compendium Elucidates Complex I Disease Biology', *Cell*, 134(1), pp. 112-123.

Palikaras, K., Lionaki, E. and Tavernarakis, N. (2015) 'Coordination of mitophagy and mitochondrial biogenesis during ageing in C. elegans', *Nature*, advance online publication.

Parsons, P. and Simpson, M.V. (1967) 'Biosynthesis of DNA by isolated mitochondria: Incorporation of thymidine triphosphate-2-C14', *Science*, 155(3758), pp. 91-93.

Paul, V.D. and Lill, R. (2015) 'Biogenesis of cytosolic and nuclear iron–sulfur proteins and their role in genome stability', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853(6), pp. 1528-1539.

Payne, B.A.I. and Chinnery, P.F. (2015) 'Mitochondrial dysfunction in aging: Much progress but many unresolved questions', *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1847(11), pp. 1347-1353.

Payne, B.A.I., Wilson, I.J., Hateley, C.A., Horvath, R., Santibanez-Koref, M., Samuels, D.C., Price, D.A. and Chinnery, P.F. (2011) 'Mitochondrial aging is accelerated by anti-retroviral therapy through the clonal expansion of mtDNA mutations', *Nature Genetics*, 43(8), pp. 806-810.

Payne, B.A.I., Wilson, I.J., Yu-Wai-Man, P., Coxhead, J., Deehan, D., Horvath, R., Taylor, R.W., Samuels, D.C., Santibanez-Koref, M. and Chinnery, P.F. (2013) 'Universal heteroplasmy of human mitochondrial DNA', *Human Molecular Genetics*, 22(2), pp. 384-390.

Pe'er, I., Yelensky, R., Altshuler, D. and Daly, M.J. (2008) 'Estimation of the multiple testing burden for genomewide association studies of nearly all common variants', *Genet Epidemiol*, 32(4), pp. 381-5.

Pello, R., Martín, M.A., Carelli, V., Nijtmans, L.G., Achilli, A., Pala, M., Torroni, A., Gómez-Durán, A., Ruiz-Pesini, E., Martinuzzi, A., Smeitink, J.A., Arenas, J. and Ugalde, C. (2008) 'Mitochondrial DNA background modulates the assembly kinetics of OXPHOS complexes in a cellular model of mitochondrial disease', *Human Molecular Genetics*, 17(24), pp. 4001-4011.

Perier, C., Bender, A., Garcia-Arumi, E., Melia, M.J., Bove, J., Laub, C., Klopstock, T., Elstner, M., Mounsey, R.B., Teismann, P., Prolla, T., Andreu, A.L. and Vila, M. (2013a) 'Accumulation of mitochondrial DNA deletions within dopaminergic neurons triggers neuroprotective mechanisms', *Brain*, 136(Pt 8), pp. 2369-78.

Perier, C., Bender, A., García-Arumí, E., Melià, M.J., Bové, J., Laub, C., Klopstock, T., Elstner, M., Mounsey, R.B., Teismann, P., Prolla, T., Andreu, A.L. and Vila, M.

(2013b) 'Accumulation of mitochondrial DNA deletions within dopaminergic neurons triggers neuroprotective mechanisms', *Brain*, 136(8), pp. 2369-2378.

Petukhova, L., Duvic, M., Hordinsky, M., Norris, D., Price, V., Shimomura, Y., Kim, H., Singh, P., Lee, A., Chen, W.V., Meyer, K.C., Paus, R., Jahoda, C.A., Amos, C.I., Gregersen, P.K. and Christiano, A.M. (2010) 'Genome-wide association study in alopecia areata implicates both innate and adaptive immunity', *Nature*, 466(7302), pp. 113-7.

Pfeffer, S.R. (2001) 'Rab GTPases: specifying and deciphering organelle identity and function', *Trends in cell biology*, 11(12), pp. 487-491.

Phillips, N.R., Sprouse, M.L. and Roby, R.K. (2014) 'Simultaneous quantification of mitochondrial DNA copy number and deletion ratio: A multiplex real-time PCR assay', *Scientific Reports*, 4.

Picard, M., McManus, M.J., Csordás, G., Várnai, P., Dorn Ii, G.W., Williams, D., Hajnóczky, G. and Wallace, D.C. (2015) 'Trans-mitochondrial coordination of cristae at regulated membrane junctions', *Nat Commun*, 6.

Pickrell, A.M. and Youle, R.J. (2015) 'The Roles of PINK1, Parkin and Mitochondrial Fidelity in Parkinson's Disease', *Neuron*, 85(2), pp. 257-273.

Pierron, D., Letellier, T. and Grossman, L.I. (2012) 'Mitogroup: Continent-specific clusters of mitochondrial OXPHOS complexes based on nuclear non-synonymous polymorphisms', *Mitochondrion*, 12(2), pp. 237-241.

Pisano, A., Preziuso, C., Iommarini, L., Perli, E., Grazioli, P., Campese, A.F., Maresca, A., Montopoli, M., Masuelli, L., Sadun, A.A., d'Amati, G., Carelli, V., Ghelli, A. and Giordano, C. (2015) 'Targeting estrogen receptor β as preventive therapeutic strategy for Leber's hereditary optic neuropathy', *Human Molecular Genetics*, 24(24), pp. 6921-6931.

Pitceathly, R.D.S., Rahman, S. and Hanna, M.G. (2012) 'Single deletions in mitochondrial DNA – Molecular mechanisms and disease phenotypes in clinical practice', *Neuromuscular Disorders*, 22(7), pp. 577-586.

Pittis, A.A. and Gabaldón, T. (2016) 'Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry', *Nature*, 531(7592), pp. 101-104.

Politi, Y., Gal, L., Kalifa, Y., Ravid, L., Elazar, Z. and Arama, E. (2014) 'Paternal mitochondrial destruction after fertilization is mediated by a common endocytic and autophagic pathway in drosophila', *Developmental Cell*, 29(3), pp. 305-320.

Polyak, K., Li, Y., Zhu, H., Lengauer, C., Willson, J.K. and Markowitz, S.D. (1998) 'Somatic mutations of the mitochondrial genome in human colorectal tumours', *Nat Genet*, 20.

Ponti, M., Forrow, S.M., Souhami, R.L., D'Incalci, M. and Hartley, J.A. (1991) 'Measurement of the sequence specificity of covalent DNA modification by antineoplastic agents using Taq DNA polymerase', *Nucleic Acids Research*, 19(11), pp. 2929-2933.

Poole, O.V., Hanna, M.G. and Pitceathly, R.D.S. (2015) 'Mitochondrial disorders: Disease mechanisms and therapeutic approaches', *Discovery Medicine*, 20(111), pp. 325-331.

Posse, V. and Gustafsson, C.M. (2017) 'Human Mitochondrial Transcription Factor B2 Is Required for Promoter Melting during Initiation of Transcription', *Journal of Biological Chemistry*, 292(7), pp. 2637-2645.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) 'Principal components analysis corrects for stratification in genome-wide association studies', *Nat Genet*, 38(8), pp. 904-909.

Pringsheim, T., Jette, N., Frolkis, A. and Steeves, T.D.L. (2014) 'The prevalence of Parkinson's disease: A systematic review and meta-analysis', *Movement Disorders*, 29(13), pp. 1583-1590.

Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) 'LocusZoom: regional visualization of genome-wide association scan results', *Bioinformatics*, 26(18), pp. 2336-7.

Puigserver, P. and Spiegelman, B.M. (2003) 'Peroxisome Proliferator-Activated Receptor-γ Coactivator 1α (PGC-1α): Transcriptional Coactivator and Metabolic Regulator', *Endocrine Reviews*, 24(1), pp. 78-90.

Puomila, A., Hamalainen, P., Kivioja, S., Savontaus, M.-L., Koivumaki, S., Huoponen, K. and Nikoskelainen, E. (2007) 'Epidemiology and penetrance of Leber hereditary optic neuropathy in Finland', *Eur J Hum Genet*, 15(10), pp. 1079-1089.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ and Sham PC (2007) 'PLINK 1.9'. Available at: http://pngu.mgh.harvard.edu/purcell/plink/.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J. and Sham, P.C. (2007) 'PLINK: A tool set for whole-genome association and population-based linkage analyses', *American Journal of Human Genetics*, 81(3), pp. 559-575.

Pyakurel, A., Savoia, C., Hess, D. and Scorrano, L. (2014) 'Extracellular Regulated Kinase Phosphorylates Mitofusin 1 to Control Mitochondrial Morphology and Apoptosis', *Molecular Cell*.

Pye, D., Kyriakouli, D.S., Taylor, G.A., Johnson, R., Elstner, M., Meunier, B., Chrzanowska-Lightowlers, Z.M.A., Taylor, R.W., Turnbull, D.M. and Lightowlers, R.N. (2006) 'Production of transmitochondrial cybrids containing naturally occurring pathogenic mtDNA variants', *Nucleic Acids Research*, 34(13).

Pyle, A., Anugrha, H., Kurzawa-Akanbi, M., Yarnall, A., Burn, D. and Hudson, G. (2015a) 'Reduced mitochondrial DNA copy number is a biomarker of Parkinson's disease', *Neurobiology of Aging*, 38, pp. 216.e7-216.e10.

Pyle, A., Brennan, R., Kurzawa-Akanbi, M., Yarnall, A., Thouin, A., Mollenhauer, B., Burn, D., Chinnery, P.F. and Hudson, G. (2015b) 'Reduced cerebrospinal fluid mitochondrial DNA is a biomarker for early-stage Parkinson's disease', *Annals of Neurology*, 78(6), pp. 1000-1004.

Pyle, A., Burn, D., Gordon, C., Swan, C., Chinnery, P. and Baudouin, S. (2010) 'Fall in circulating mononuclear cell mitochondrial DNA content in human sepsis', *Intensive Care Medicine*, 36(6), pp. 956-962.

Pyle, A., Foltynie, T., Tiangyou, W., Lambert, C., Keers, S.M., Allcock, L.M., Davison, J., Lewis, S.J., Perry, R.H., Barker, R., Burn, D.J. and Chinnery, P.F. (2005) 'Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD', *Annals of Neurology*, 57(4), pp. 564-567.

Pyle, A., Lowes, H., Brennan, R., Kurzawa-Akanbi, M., Yarnall, A., Burn, D. and Hudson, G. (2016) 'Reduced mitochondrial DNA is not a biomarker of depression in Parkinson's disease', *Movement Disorders*, 31(12), pp. 1923-1924.

Pyle, A., Taylor, R.W., Durham, S.E., Deschauer, M., Schaefer, A.M., Samuels, D.C. and Chinnery, P.F. (2007) 'Depletion of mitochondrial DNA in leucocytes harbouring the 3243A→G mtDNA mutation', *Journal of Medical Genetics*, 44(1), pp. 69-74.

Qi, Z. and Ding, S. (2012) 'Transcriptional Regulation by Nuclear Corepressors and PGC-1α: Implications for Mitochondrial Quality Control and Insulin Sensitivity', *PPAR Research*, 2012, p. 12.

Quinlan, A.R. and Hall, I.M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841-842.

Quiros, P.M., Mottis, A. and Auwerx, J. (2016) 'Mitonuclear communication in homeostasis and stress', *Nat Rev Mol Cell Biol*, 17(4), pp. 213-226.

Rackham, O., Shearwood, A.-M.J., Mercer, T.R., Davies, S.M.K., Mattick, J.S. and Filipovska, A. (2011) 'Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins', *RNA*, 17(12), pp. 2085-2093.

Radzvilavicius, A.L., Hadjivasiliou, Z., Pomiankowski, A. and Lane, N. (2016) 'Selection for Mitochondrial Quality Drives Evolution of the Germline', *PLoS Biology*, 14(12).

Rajala, N., Gerhold, J.M., Martinsson, P., Klymov, A. and Spelbrink, J.N. (2014) 'Replication factors transiently associate with mtDNA at the mitochondrial inner membrane to facilitate replication', *Nucleic Acids Research*, 42(2), pp. 952-967.

Ramachandran, A., Basu, U., Sultana, S., Nandakumar, D. and Patel, S.S. (2017) 'Human mitochondrial transcription factors TFAM and TFB2M work synergistically in promoter melting during transcription initiation', *Nucleic Acids Research*, 45(2), pp. 861-874.

Ramanathan, A. and Schreiber, S.L. (2009) 'Direct control of mitochondrial function by mTOR', *Proceedings of the National Academy of Sciences*, 106(52), pp. 22229-22232.

Randle, P.J., Garland, P.B., Hales, C.N. and Newsholme, E.A. (1963) 'The glucose fatty-acid cycle its role in insulin sensitivity and the metabolic disturbances of diabetes mellitus', *The Lancet*, 281(7285), pp. 785-789.

Rassat, J., Robenek, H. and Themann, H. (1981) 'Structural relationship between desmosomes and mitochondria in human livers exhibiting a wide range of diseases', *The American Journal of Pathology*, 105(3), pp. 207-211.

Rebouças, E.d.L., Costa, J.J.d.N., Passos, M.J., Passos, J.R.d.S., Hurk, R.v.d. and Silva, J.R.V. (2013) 'Real time PCR and importance of housekeepings genes for normalization and quantification of mRNA expression in different tissues', *Brazilian Archives of Biology and Technology*, 56, pp. 143-154.

Reddy, P., Ocampo, A., Suzuki, K., Luo, J., Bacman, Sandra R., Williams, Sion L., Sugawara, A., Okamura, D., Tsunekawa, Y., Wu, J., Lam, D., Xiong, X., Montserrat, N., Esteban, Concepcion R., Liu, G.-H., Sancho-Martinez, I., Manau, D., Civico, S., Cardellach, F., del Mar O'Callaghan, M., Campistol, J., Zhao, H., Campistol, Josep M., Moraes, Carlos T. and Izpisua Belmonte, Juan C. (2015) 'Selective Elimination of Mitochondrial Mutations in the Germline by Genome Editing', *Cell*, 161(3), pp. 459-469.

Reinecke, F., Smeitink, J.A.M. and van der Westhuizen, F.H. (2009) 'OXPHOS gene expression and control in mitochondrial disorders', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1792(12), pp. 1113-1121.

Relton, C.L., Gaunt, T., McArdle, W., Ho, K., Duggirala, A., Shihab, H., Woodward, G., Lyttleton, O., Evans, D.M., Reik, W., Paul, Y.L., Ficz, G., Ozanne, S.E., Wipat, A., Flanagan, K., Lister, A., Heijmans, B.T., Ring, S.M. and Davey Smith, G. (2015) 'Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES)', *Int J Epidemiol*, 44(4), pp. 1181-90.

Ren, M., Phoon, C.K.L. and Schlame, M. (2014) 'Metabolism and function of mitochondrial cardiolipin', *Progress in Lipid Research*, 55(1), pp. 1-16.

Replication, D.I.G., Meta-analysis, C., Asian Genetic Epidemiology Network Type 2 Diabetes, C., South Asian Type 2 Diabetes, C., Mexican American Type 2 Diabetes, C., Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples, C., Mahajan, A., Go, M.J., Zhang, W., Below, J.E., Gaulton, K.J.,

Ferreira, T., Horikoshi, M., Johnson, A.D., Ng, M.C., Prokopenko, I., Saleheen, D., Wang, X., Zeggini, E., Abecasis, G.R., Adair, L.S., Almgren, P., Atalay, M., Aung, T., Baldassarre, D., Balkau, B., Bao, Y., Barnett, A.H., Barroso, I., Basit, A., Been, L.F., Beilby, J., Bell, G.I., Benediktsson, R., Bergman, R.N., Boehm, B.O., Boerwinkle, E., Bonnycastle, L.L., Burtt, N., Cai, Q., Campbell, H., Carey, J., Cauchi, S., Caulfield, M., Chan, J.C., Chang, L.C., Chang, T.J., Chang, Y.C., Charpentier, G., Chen, C.H., Chen, H., Chen, Y.T., Chia, K.S., Chidambaram, M., Chines, P.S., Cho, N.H., Cho, Y.M., Chuang, L.M., Collins, F.S., Cornelis, M.C., Couper, D.J., Crenshaw, A.T., van Dam, R.M., Danesh, J., Das, D., de Faire, U., Dedoussis, G., Deloukas, P., Dimas, A.S., Dina, C., Doney, A.S., Donnelly, P.J., Dorkhan, M., van Duijn, C., Dupuis, J., Edkins, S., Elliott, P., Emilsson, V., Erbel, R., Eriksson, J.G., Escobedo, J., Esko, T., Eury, E., Florez, J.C., Fontanillas, P., Forouhi, N.G., Forsen, T., Fox, C., Fraser, R.M., Frayling, T.M., Froguel, P., Frossard, P., Gao, Y., Gertow, K., Gieger, C., Gigante, B., Grallert, H., Grant, G.B., Grrop, L.C., Groves, C.J., et al. (2014) 'Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility', *Nat Genet*, 46(3), pp. 234-44.

Reznik, E., Miller, M.L., Şenbabaoğlu, Y., Riaz, N., Sarungbam, J., Tickoo, S.K., Al-Ahmadie, H.A., Lee, W., Seshan, V.E., Hakimi, A.A. and Sander, C. (2016) 'Mitochondrial DNA copy number variation across human cancers', *eLife*, 5, p. e10769.

Rice, A.C., Keeney, P.M., Algarzae, N.K., Ladd, A.C., Thomas, R.R. and Bennett Jr, J.P. (2014) 'Mitochondrial DNA copy numbers in pyramidal neurons are decreased and mitochondrial biogenesis transcriptome signaling is disrupted in Alzheimer's disease hippocampi', *Journal of Alzheimer's Disease*, 40(2), pp. 319-330.

Richmond, R.C., Simpkin, A.J., Woodward, G., Gaunt, T.R., Lyttleton, O., McArdle, W.L., Ring, S.M., Smith, A.D., Timpson, N.J., Tilling, K., Davey Smith, G. and Relton, C.L. (2015) 'Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC)', *Hum Mol Genet*, 24(8), pp. 2201-17.

Richter, C., Park, J.W. and Ames, B.N. (1988) 'Normal oxidative damage to mitochondrial and nuclear DNA is extensive', *Proceedings of the National Academy of Sciences of the United States of America*, 85(17), pp. 6465-6467.

Ridge, P.G., Maxwell, T.J., Foutz, S.J., Bailey, M.H., Corcoran, C.D., Tschanz, J.T., Norton, M.C., Munger, R.G., O'Brien, E., Kerber, R.A., Cawthon, R.M. and Kauwe, J.S.K. (2014) 'Mitochondrial genomic variation associated with higher mitochondrial copy number: the Cache County Study on Memory Health and Aging', *BMC Bioinformatics*, 15(Suppl 7), pp. S6-S6.

Roberts, R.F., Tang, M.Y., Fon, E.A. and Durcan, T.M. (2016) 'Defending the mitochondria: The pathways of mitophagy and mitochondrial-derived vesicles', *Int J Biochem Cell Biol*, 79, pp. 427-436.

Robin, E.D. and Wong, R. (1988) 'Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells', *J Cell Physiol*, 136.

Robinson, B.H. (2006) 'Lactic acidemia and mitochondrial disease', *Molecular Genetics and Metabolism*, 89(1-2), pp. 3-13.

Rocha, M.C., Grady, J.P., Grünewald, A., Vincent, A., Dobson, P.F., Taylor, R.W., Turnbull, D.M. and Rygiel, K.A. (2015) 'A novel immunofluorescent assay to investigate oxidative phosphorylation deficiency in mitochondrial myopathy: understanding mechanisms and improving diagnosis', 5, p. 15037.

Röder, B., Frühwirth, K., Vogl, C., Wagner, M. and Rossmanith, P. (2010) 'Impact of Long-Term Storage on Stability of Standard DNA for Nucleic Acid-Based Methods', *Journal of Clinical Microbiology*, 48(11), pp. 4260-4262.

Rodgers, J.T. and Puigserver, P. (2007) 'Fasting-dependent glucose and lipid metabolic response through hepatic sirtuin 1', *Proceedings of the National Academy of Sciences*, 104(31), pp. 12861-12866.

Rodriguez-Enriquez, S., Kai, Y., Maldonado, E., Currin, R.T. and Lemasters, J.J. (2009) 'Roles of mitophagy and the mitochondrial permeability transition in remodeling of cultured rat hepatocytes', *Autophagy*, 5(8), pp. 1099-1106.

Rohland, N., Siedel, H. and Hofreiter, M. (2010) 'A rapid column-based ancient DNA extraction method for increased sample throughput', *Molecular Ecology Resources*, 10(4), pp. 677-683.

Rooney, J.P., Ryde, I.T., Sanders, L.H., Howlett, E.H., Colton, M.D., Germ, K.E., Mayer, G.D., Greenamyre, J.T. and Meyer, J.N. (2015) 'PCR Based Determination of

Mitochondrial DNA Copy Number in Multiple Species', *Methods in molecular biology (Clifton, N.J.)*, 1241, pp. 23-38.

Ropp, P.A. and Copeland, W.C. (1996) 'Cloning and characterization of the human mitochondrial DNA polymerase, DNA polymerase γ', *Genomics*, 36(3), pp. 449-458.

Rose, G., Passarino, G., Carrieri, G., Altomare, K., Greco, V., Bertolini, S., Bonafè, M., Franceschi, C. and De Benedictis, G. (2001) 'Paradoxes in longevity: Sequence analysis of mtDNA haplogroup J in centenarians', *European Journal of Human Genetics*, 9(9), pp. 701-707.

Ross, K.S., Haites, N.E. and Kelly, K.F. (1990) 'Repeated freezing and thawing of peripheral blood and DNA in suspension: Effects on DNA yield and integrity', *Journal of Medical Genetics*, 27(9), pp. 569-570.

Ross, O.A., McCormack, R., Curran, M.D., Duguid, R.A., Barnett, Y.A., Rea, I.M. and Middleton, D. (2001) 'Mitochondrial DNA polymorphism: its role in longevity of the Irish population', *Exp Gerontol*, 36(7), pp. 1161-78.

Rossignol, R., Faustin, B., Rocher, C., Malgat, M., Mazat, J.P. and Letellier, T. (2003) 'Mitochondrial threshold effects', *Biochem J*, 370(Pt 3), pp. 751-62.

Rossignol, R., Malgat, M., Mazat, J.-P. and Letellier, T. (1999) 'Threshold Effect and Tissue Specificity: Implication for Mitochondrial Cytopathies', *Journal of Biological Chemistry*, 274(47), pp. 33426-33432.

Rossouw, J.E., Anderson, G.L., Prentice, R.L., LaCroix, A.Z., Kooperberg, C., Stefanick, M.L., Jackson, R.D., Beresford, S.A.A., Howard, B.V., Johnson, K.C., Kotchen, J.M. and Ockene, J. (2002) 'Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the women's health initiative randomized controlled trial', *Journal of the American Medical Association*, 288(3), pp. 321-333.

Rötig, A. and Poulton, J. (2009) 'Genetic causes of mitochondrial DNA depletion in humans', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1792(12), pp. 1103-1108.

Rouault, T.A. and Tong, W.-H. (2005) 'Iron-sulphur cluster biogenesis and mitochondrial iron homeostasis', *Nat Rev Mol Cell Biol*, 6(4), pp. 345-351.

Rucker, G., Carpenter, J.R. and Schwarzer, G. (2011) 'Detecting and adjusting for small-study effects in meta-analysis', *Biom J*, 53(2), pp. 351-68.

Ruhanen, H., Borrie, S., Szabadkai, G., Tyynismaa, H., Jones, A.W.E., Kang, D., Taanman, J.-W. and Yasukawa, T. (2010) 'Mitochondrial single-stranded DNA binding protein is required for maintenance of mitochondrial DNA and 7S DNA but is not required for mitochondrial nucleoid organisation', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1803(8), pp. 931-939.

Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V. and Wallace, D.C. (2004) 'Effects of purifying and adaptive selection on regional variation in human mtDNA', *Science*, 303(5655), pp. 223-6.

Rushton, D.H., Dover, R., Sainsbury, A.W., Norris, M.J., Gilkes, J.J.H. and Ramsay, I.D. (2001) 'Why should women have lower reference limits for haemoglobin and ferritin concentrations than men?', *BMJ : British Medical Journal*, 322(7298), pp. 1355-1357.

Russell, O. and Turnbull, D. (2014) 'Mitochondrial DNA disease-molecular insights and potential routes to a cure', *Experimental Cell Research*, 325(1), pp. 38-43.

Rygiel, K.A., Grady, J.P., Taylor, R.W., Tuppen, H.A.L. and Turnbull, D.M. (2015) 'Triplex real-time PCR–an improved method to detect a wide spectrum of mitochondrial DNA deletions in single cells', *Scientific Reports*, 5, p. 9906.

Rygiel, K.A., Tuppen, H.A., Grady, J.P., Vincent, A., Blakely, E.L., Reeve, A.K., Taylor, R.W., Picard, M., Miller, J. and Turnbull, D.M. (2016) 'Complex mitochondrial DNA rearrangements in individual cells from patients with sporadic inclusion body myositis', *Nucleic Acids Research*.

Saada, A. (2004) 'Deoxyribonucleotides and disorders of mitochondrial DNA integrity', *DNA and Cell Biology*, 23(12), pp. 797-806.

Sack, M.N. and Finkel, T. (2012) 'Mitochondrial metabolism, sirtuins, and aging', *Cold Spring Harbor Perspectives in Biology*, 4(12).

Sadatomi, D., Nakashioya, K., Mamiya, S., Honda, S., Kameyama, Y., Yamamura, Y., Tanimura, S. and Takeda, K. (2017) 'Mitochondrial function is required for

extracellular ATP-induced NLRP3 inflammasome activation', *The Journal of Biochemistry*, 161(6), pp. 503-512.

Sampson, T.R., Debelius, J.W., Thron, T., Janssen, S., Shastri, G.G., Ilhan, Z.E., Challis, C., Schretter, C.E., Rocha, S., Gradinaru, V., Chesselet, M.-F., Keshavarzian, A., Shannon, K.M., Krajmalnik-Brown, R., Wittung-Stafshede, P., Knight, R. and Mazmanian, S.K. (2016) 'Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinsons Disease', *Cell*, 167(6), pp. 1469-1480.e12.

Samuels, D.C., Li, C., Li, B., Song, Z., Torstenson, E., Boyd Clay, H., Rokas, A., Thornton-Wells, T.A., Moore, J.H., Hughes, T.M., Hoffman, R.D., Haines, J.L., Murdock, D.G., Mortlock, D.P. and Williams, S.M. (2013) 'Recurrent tissue-specific mtDNA mutations are common in humans', *PLoS Genet*, 9(11), p. e1003929.

Santen, R.J., Allred, D.C., Ardoin, S.P., Archer, D.F., Boyd, N., Braunstein, G.D., Burger, H.G., Colditz, G.A., Davis, S.R., Gambacciani, M., Gower, B.A., Henderson, V.W., Jarjour, W.N., Karas, R.H., Kleerekoper, M., Lobo, R.A., Manson, J.E., Marsden, J., Martin, K.A., Martin, L., Pinkerton, J.V., Rubinow, D.R., Teede, H., Thiboutot, D.M. and Utian, W.H. (2010) 'Postmenopausal Hormone Therapy: An Endocrine Society Scientific Statement', *The Journal of Clinical Endocrinology & Metabolism*, 95(7_supplement_1), pp. s1-s66.

Santorelli, F.M., Shanske, S., Macaya, A., DeVivo, D.C. and DiMauro, S. (1993) 'The mutation at nt 8993 of mitochondrial DNA is a common cause of Leigh's syndrome', *Annals of Neurology*, 34(6), pp. 827-834.

Santoro, A., Balbi, V., Balducci, E., Pirazzini, C., Rosini, F., Tavano, F., Achilli, A., Siviero, P., Minicuci, N., Bellavista, E., Mishto, M., Salvioli, S., Marchegiani, F., Cardelli, M., Olivieri, F., Nacmias, B., Chiamenti, A.M., Benussi, L., Ghidoni, R., Rose, G., Gabelli, C., Binetti, G., Sorbi, S., Crepaldi, G., Passarino, G., Torroni, A. and Franceschi, C. (2010) 'Evidence for sub-haplogroup h5 of mitochondrial DNA as a risk factor for late onset Alzheimer's disease', *PLoS One*, 5(8), p. e12037.

Sas, K.M., Kayampilly, P., Byun, J., Nair, V., Hinder, L.M., Hur, J., Zhang, H., Lin, C., Qi, N.R., Michailidis, G., Groop, P.-H., Nelson, R.G., Darshi, M., Sharma, K., Schelling, J.R., Sedor, J.R., Pop-Busui, R., Weinberg, J.M., Soleimanpour, S.A., Abcouwer, S.F., Gardner, T.W., Burant, C.F., Feldman, E.L., Kretzler, M., Brosius,

F.C., III and Pennathur, S. (2016) 'Tissue-specific metabolic reprogramming drives nutrient flux in diabetic complications', *JCI Insight*, 1(15).

Sato, M. and Sato, K. (2011) 'Degradation of paternal mitochondria by fertilization-triggered autophagy in C. elegans embryos', *Science*, 334(6059), pp. 1141-1144.

Saxena, R., de Bakker, P.I., Singer, K., Mootha, V., Burtt, N., Hirschhorn, J.N., Gaudet, D., Isomaa, B., Daly, M.J., Groop, L., Ardlie, K.G. and Altshuler, D. (2006) 'Comprehensive association testing of common mitochondrial DNA variation in metabolic disease', *Am J Hum Genet*, 79(1), pp. 54-61.

Saxonov, S., Berg, P. and Brutlag, D.L. (2006) 'A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters', *Proceedings of the National Academy of Sciences of the United States of America*, 103(5), pp. 1412-1417.

Scarpulla, R.C. (2006) 'Nuclear control of respiratory gene expression in mammalian cells', *Journal of Cellular Biochemistry*, 97(4), pp. 673-683.

Scarpulla, R.C. (2011) 'Metabolic control of mitochondrial biogenesis through the PGC-1 family regulatory network', *Biochimica et biophysica acta*, 1813(7), pp. 1269-1278.

Schagger, H. (2006) 'Tricine-SDS-PAGE', *Nat. Protocols*, 1(1), pp. 16-22.

Schatz, G. (1995) 'Mitochondria: beyond oxidative phosphorylation', *BBA - Molecular Basis of Disease*, 1271(1), pp. 123-126.

Schaudien, D., Baumgärtner, W. and Herden, C. (2007) 'High Preservation of DNA Standards Diluted in 50% Glycerol', *Diagnostic Molecular Pathology*, 16(3), pp. 153-157.

Schenkel, L.C. and Bakovic, M. (2014) 'Formation and regulation of mitochondrial membranes', *International Journal of Cell Biology*.

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014) 'Biological insights from 108 schizophrenia-associated genetic loci', *Nature*, 511(7510), pp. 421-427.

Schmidt, H., Bastholt, L., Geertsen, P., Christensen, I.J., Larsen, S., Gehl, J. and von der Maase, H. (2005) 'Elevated neutrophil and monocyte counts in peripheral blood are associated with poor survival in patients with metastatic melanoma: a prognostic model', *British Journal of Cancer*, 93(3), pp. 273-278.

Schneider, W.C. and Kuff, E.L. (1965) 'The isolation and some properties of rat liver mitochondrial deoxyribonucleic acid', *Proceedings of the National Academy of Sciences of the United States of America*, 54(6), pp. 1650-1658.

Schultz, R.A., Swoap, S.J., McDaniel, L.D., Zhangt, B., Koon, E.C., Garry, D.J., Li, K. and Williams, R.S. (1998) 'Differential expression of mitochondrial DNA replication factors in mammalian tissues', *Journal of Biological Chemistry*, 273(6), pp. 3447-3451.

Sciacco, M., Bonilla, E., Schon, E.A., Dimauro, S. and Moraes, C.T. (1994) 'Distribution of wild-type and common deletion forms of mtDNA in normal and respiration-deficient muscle fibers from patients with mitochondrial myopathy', *Human Molecular Genetics*, 3(1), pp. 13-19.

Seidel-Rogol, B.L. and Shadel, G.S. (2002) 'Modulation of mitochondrial transcription in response to mtDNA depletion and repletion in HeLa cells', *Nucleic Acids Research*, 30(9), pp. 1929-1934.

Seifer, D.B., DeJesus, V. and Hubbard, K. (2002) 'Mitochondrial deletions in luteinized granulosa cells as a function of age in women undergoing in vitro fertilization', *Fertility and Sterility*, 78(5), pp. 1046-1048.

Selak, M.A., Lyver, E., Micklow, E., Deutsch, E.C., Önder, Ö., Selamoglu, N., Yager, C., Knight, S., Carroll, M., Daldal, F., Dancis, A., Lynch, D.R. and Sarry, J.-E. (2011) 'Blood cells from Friedreich ataxia patients harbor frataxin deficiency without a loss of mitochondrial function', *Mitochondrion*, 11(2), pp. 342-350.

Shao, W., Khin, S. and Kopp, W.C. (2012) 'Characterization of effect of repeated freeze and thaw cycles on stability of genomic DNA using pulsed field gel electrophoresis', *Biopreservation and Biobanking*, 10(1), pp. 4-11.

Shaughnessy, D.T., McAllister, K., Worth, L., Haugen, A.C., Meyer, J.N., Domann, F.E., Van Houten, B., Mostoslavsky, R., Bultman, S.J., Baccarelli, A.A., Begley, T.J., Sobol, R.W., Hirschey, M.D., Ideker, T., Santos, J.H., Copeland, W.C., Tice, R.R.,

Balshaw, D.M. and Tyson, F.L. (2014) 'Mitochondria, energetics, epigenetics, and cellular responses to stress', *Environmental Health Perspectives*, 122(12), pp. 1271-8.

Shen, J., Platek, M., Mahasneh, A., Ambrosone, C.B. and Zhao, H. (2010) 'Mitochondrial copy number and risk of breast cancer: A pilot study', *Mitochondrion*, 10(1), pp. 62-68.

Shen, J., Song, R., Lu, Z. and Zhao, H. (2016) 'Mitochondrial DNA copy number in whole blood and glioma risk: A case control study', *Molecular Carcinogenesis*, 55(12), pp. 2089-2094.

Shen, L., Wei, J., Chen, T., He, J., Qu, J., He, X., Jiang, L., Qu, Y., Fang, H., Chen, G., Lu, J. and Bai, Y. (2011) 'Evaluating mitochondrial DNA in patients with breast cancer and benign breast disease', *J Cancer Res Clin Oncol*, 137(4), pp. 669-75.

Sherman, B.M., West, J.H. and Korenman, S.G. (1976) 'The Menopausal Transition: Analysis of LH, FSH, Estradiol, and Progesterone Concentrations During Menstrual Cycles of Older Women', *The Journal of Clinical Endocrinology & Metabolism*, 42(4), pp. 629-636.

Shin, J. and Lee, C. (2015) 'Statistical power for identifying nucleotide markers associated with quantitative traits in genome-wide association analysis using a mixed model', *Genomics*, 105(1), pp. 1-4.

Shoffner, J.M., Brown, M.D., Torroni, A., Lott, M.T., Cabell, M.F., Mirra, S.S., Beal, M.F., Yang, C.C., Gearing, M., Salvo, R., Watts, R.L., Juncos, J.L., Hansen, L.A., Crain, B.J., Fayad, M., Reckord, C.L. and Wallace, D.C. (1993) 'Mitochondrial DNA Variants Observed in Alzheimer Disease and Parkinson Disease Patients', *Genomics*, 17(1), pp. 171-184.

Shoffner, J.M., Lott, M.T., Voljavec, A.S., Soueidan, S.A., Costigan, D.A. and Wallace, D.C. (1989) 'Spontaneous Kearns-Sayre/chronic external ophthalmoplegia plus syndrome associated with a mitochondrial DNA deletion: A slip-replication model and metabolic therapy', *Proceedings of the National Academy of Sciences of the United States of America*, 86(20), pp. 7952-7956.

Shoubridge, E.A. and Wai, T. 77 (2007) 'Mitochondrial DNA and the Mammalian Oocyte' *Current Topics in Developmental Biology*. pp. 87-111.

Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Magi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., Workalemahu, T., Wu, J.M.W., Buchkovich, M.L., Heard-Costa, N.L., Roman, T.S., Drong, A.W., Song, C., Gustafsson, S., Day, F.R., Esko, T., Fall, T., Kutalik, Z., Luan, J.a., Randall, J.C., Scherag, A., Vedantam, S., Wood, A.R., Chen, J., Fehrmann, R., Karjalainen, J., Kahali, B., Liu, C.-T., Schmidt, E.M., Absher, D., Amin, N., Anderson, D., Beekman, M., Bragg-Gresham, J.L., Buyske, S., Demirkan, A., Ehret, G.B., Feitosa, M.F., Goel, A., Jackson, A.U., Johnson, T., Kleber, M.E., Kristiansson, K., Mangino, M., Mateo Leach, I., Medina-Gomez, C., Palmer, C.D., Pasko, D., Pechlivanis, S., Peters, M.J., Prokopenko, I., Stancakova, A., Ju Sung, Y., Tanaka, T., Teumer, A., Van Vliet-Ostaptchouk, J.V., Yengo, L., Zhang, W., Albrecht, E., Arnlov, J., Arscott, G.M., Bandinelli, S., Barrett, A., Bellis, C., Bennett, A.J., Berne, C., Bluher, M., Bohringer, S., Bonnet, F., Bottcher, Y., Bruinenberg, M., Carba, D.B., Caspersen, I.H., Clarke, R., Warwick Daw, E., Deelen, J., Deelman, E., Delgado, G., Doney, A.S.F., Eklund, N., Erdos, M.R., Estrada, K., Eury, E., Friedrich, N., Garcia, M.E., Giedraitis, V., Gigante, B., Go, A.S., Golay, A., Grallert, H., Grammer, T.B., Graszler, J., Grewal, J., Groves, C.J., Haller, T., Hallmans, G., et al. (2015) 'New genetic loci link adipose and insulin biology to body fat distribution', *Nature*, 518(7538), pp. 187-196.

Sidore, C., Busonero, F., Maschio, A., Porcu, E., Naitza, S., Zoledziewska, M., Mulas, A., Pistis, G., Steri, M., Danjou, F., Kwong, A., Ortega Del Vecchyo, V.D., Chiang, C.W.K., Bragg-Gresham, J., Pitzalis, M., Nagaraja, R., Tarrier, B., Brennan, C., Uzzau, S., Fuchsberger, C., Atzeni, R., Reinier, F., Berutti, R., Huang, J., Timpson, N.J., Toniolo, D., Gasparini, P., Malerba, G., Dedoussis, G., Zeggini, E., Soranzo, N., Jones, C., Lyons, R., Angius, A., Kang, H.M., Novembre, J., Sanna, S., Schlessinger, D., Cucca, F. and Abecasis, G.R. (2015) 'Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers', *Nat Genet*, 47(11), pp. 1272-1281.

Silva, J.P., Köhler, M., Graff, C., Oldfors, A., Magnuson, M.A., Berggren, P.O. and Larsson, N.G. (2000) 'Impaired insulin secretion and β-cell loss in tissue-specific knockout mice with mitochondrial diabetes', *Nature Genetics*, 26(3), pp. 336-340.

Simons, K. and Zerial, M. (1993) 'Rab proteins and the road maps for intracellular transport', *Neuron*, 11(5), pp. 789-799.

Singal, R. and Ginder, G.D. (1999) 'DNA Methylation', *Blood*, 93(12), pp. 4059-4070.

Singh, K.K. (2004) 'Mitochondrial dysfunction is a common phenotype in aging and cancer', *Ann N Y Acad Sci*, 1019.

Skladal, D., Halliday, J. and Thorburn, D.R. (2003) 'Minimum birth prevalence of mitochondrial respiratory chain disorders in children', *Brain*, 126(8), pp. 1905-1912.

Skulachev, V.P. (2001) 'Mitochondrial filaments and clusters as intracellular power-transmitting cables', *Trends in Biochemical Sciences*, 26(1), pp. 23-29.

Smeitink, J., van den Heuvel, L. and DiMauro, S. (2001) 'The genetics and pathology of oxidative phosphorylation', *Nat Rev Genet*, 2(5), pp. 342-352.

Smith, A.C. and Robinson, A.J. (2011) 'A metabolic model of the mitochondrion and its use in modelling diseases of the tricarboxylic acid cycle', *BMC Systems Biology*, 5.

Soto-Hermida, A., Fernández-Moreno, M., Oreiro, N., Fernández-López, C., Rego-Pérez, I. and Blanco, F.J. (2014) 'MtDNA haplogroups and osteoarthritis in different geographic populations', *Mitochondrion*, 15(1), pp. 18-23.

Spelbrink, J.N., Li, F.Y., Tiranti, V., Nikali, K., Yuan, Q.P., Tariq, M., Wanrooij, S., Garrido, N., Comi, G., Morandi, L., Santoro, L., Toscano, A., Fabrizi, G.M., Somer, H., Croxen, R., Beeson, D., Poulton, J., Suomalainen, A., Jacobs, H.T., Zeviani, M. and Larsson, C. (2001) 'Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria', *Nature Genetics*, 28(3), pp. 223-231.

Spelbrink, J.N., Toivonen, J.M., Hakkaart, G.A.J., Kurkela, J.M., Cooper, H.M., Lehtinen, S.K., Lecrenier, N., Back, J.W., Speijer, D., Foury, F. and Jacobs, H.T. (2000) 'In vivo functional analysis of the human mitochondrial DNA polymerase POLG expressed in cultured human cells', *Journal of Biological Chemistry*, 275(32), pp. 24818-24828.

Srinivasainagendra, V., Sandel, M.W., Singh, B., Sundaresan, A., Mooga, V.P., Bajpai, P., Tiwari, H.K. and Singh, K.K. (2017) 'Migration of mitochondrial DNA in the nuclear genome of colorectal adenocarcinoma', *Genome Medicine*, 9(1), p. 31.

St. John, J.C. (2016) 'Mitochondrial DNA copy number and replication in reprogramming and differentiation', *Seminars in Cell & Developmental Biology*, 52, pp. 93-101.

StatsDirect Ltd (2013) *StatsDirect statistical software* [Computer program]. StatsDirect Ltd. Available at: http://www.statsdirect.com/.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M. and Lancet, D. (2002) 'The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses', in *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc.

Stewart, J.B. and Chinnery, P.F. (2015) 'The dynamics of mitochondrial DNA heteroplasmy: Implications for human health and disease', *Nature Reviews Genetics*, 16(9), pp. 530-542.

Stewart, J.D., Schoeler, S., Sitarz, K.S., Horvath, R., Hallmann, K., Pyle, A., Yu-Wai-Man, P., Taylor, R.W., Samuels, D.C., Kunz, W.S. and Chinnery, P.F. (2011) 'POLG mutations cause decreased mitochondrial DNA repopulation rates following induced depletion in human fibroblasts', *Biochim Biophys Acta*, 1812(3), pp. 321-5.

Stirone, C., Boroujerdi, A., Duckles, S.P. and Krause, D.N. (2005) 'Estrogen receptor activation of phosphoinositide-3 kinase, akt, and nitric oxide signaling in cerebral blood vessels: rapid and long-term effects', *Molecular pharmacology*, 67(1), pp. 105-113.

Stolk, L., Perry, J.R.B., Chasman, D.I., He, C., Mangino, M., Sulem, P., Barbalic, M., Broer, L., Byrne, E.M., Ernst, F., Esko, T., Franceschini, N., Gudbjartsson, D.F., Hottenga, J.-J., Kraft, P., McArdle, P.F., Porcu, E., Shin, S.-Y., Smith, A.V., van Wingerden, S., Zhai, G., Zhuang, W.V., Albrecht, E., Alizadeh, B.Z., Aspelund, T., Bandinelli, S., Lauc, L.B., Beckmann, J.S., Boban, M., Boerwinkle, E., Broekmans, F.J., Burri, A., Campbell, H., Chanock, S.J., Chen, C., Cornelis, M.C., Corre, T., Coviello, A.D., d'Adamo, P., Davies, G., de Faire, U., de Geus, E.J.C., Deary, I.J., Dedoussis, G.V.Z., Deloukas, P., Ebrahim, S., Eiriksdottir, G., Emilsson, V., Eriksson, J.G., Fauser, B.C.J.M., Ferreli, L., Ferrucci, L., Fischer, K., Folsom, A.R., Garcia, M.E., Gasparini, P., Gieger, C., Glazer, N., Grobbee, D.E., Hall, P., Haller, T., Hankinson, S.E., Hass, M., Hayward, C., Heath, A.C., Hofman, A., Ingelsson, E., Janssens, A.C.J.W., Johnson, A.D., Karasik, D., Kardia, S.L.R., Keyzer, J., Kiel, D.P., Kolcic, I., Kutalik, Z., Lahti, J., Lai, S., Laisk, T., Laven, J.S.E., Lawlor, D.A., Liu, J., Lopez, L.M., Louwers, Y.V., Magnusson, P.K.E., Marongiu, M., Martin, N.G., Klaric,

I.M., Masciullo, C., McKnight, B., Medland, S.E., Melzer, D., Mooser, V., Navarro, P., Newman, A.B., Nyholt, D.R., Onland-Moret, N.C., Palotie, A., Paré, G., Parker, A.N., Pedersen, N.L., et al. (2012) 'Meta-analyses identify 13 novel loci associated with age at menopause and highlights DNA repair and immune pathways', *Nature genetics*, 44(3), pp. 260-268.

Stoycheva, T., Massardo, D.R., Pesheva, M., Venkov, P., Wolf, K., Del Giudice, L. and Pontieri, P. (2007) 'Ty1 transposition induced by carcinogens in Saccharomyces cerevisiae yeast depends on mitochondrial function', *Gene*, 389(2), pp. 212-218.

Su, X. and Federoff, H.J. (2014) 'Immune Responses in Parkinson's Disease: Interplay between Central and Peripheral Immune Systems', *BioMed Research International*, 2014, p. 275178.

Suarez, J., Hu, Y., Makino, A., Fricovsky, E., Wang, H. and Dillmann, W.H. (2008) 'Alterations in mitochondrial function and cytosolic calcium induced by hyperglycemia are restored by mitochondrial transcription factor A in cardiomyocytes', *American Journal of Physiology - Cell Physiology*, 295(6), pp. C1561-C1568.

Sugiura, A., McLelland, G.-L., Fon, E.A. and McBride, H.M. (2014) 'A new pathway for mitochondrial quality control: mitochondrial-derived vesicles', *The EMBO Journal*, 33(19), pp. 2142-2156.

Suissa, S., Wang, Z., Poole, J., Wittkopp, S., Feder, J., Shutt, T.E., Wallace, D.C., Shadel, G.S. and Mishmar, D. (2009) 'Ancient mtDNA genetic variants modulate mtDNA transcription and replication', *PLoS Genetics*, 5(5).

Susin, S.A., Lorenzo, H.K., Zamzami, N., Marzo, I., Snow, B.E., Brothers, G.M., Mangion, J., Jacotot, E., Costantini, P., Loeffler, M., Larochette, N., Goodlett, D.R., Aebersold, R., Siderovski, D.P., Penninger, J.M. and Kroemer, G. (1999) 'Molecular characterization of mitochodrial apoptosis-inducing factor', *Nature*, 397(6718), pp. 441-446.

Suzuki, H., Hosokawa, Y., Nishikimi, M. and Ozawa, T. (1991) 'Existence of common homologous elements in the transcriptional regulatory regions of human nuclear genes and mitochondrial gene for the oxidative phosphorylation system', *Journal of Biological Chemistry*, 266(4), pp. 2333-2338.

Suzuki, N., Kamataki, A., Yamaki, J. and Homma, Y. (2008) 'Characterization of circulating DNA in healthy human plasma', *Clin Chim Acta*, 387(1-2), pp. 55-8.

Takamatsu, C., Umeda, S., Ohsato, T., Ohno, T., Abe, Y., Fukuoh, A., Shinagawa, H., Hamasaki, N. and Kang, D. (2002) 'Regulation of mitochondrial D-loops by transcription factor A and single-stranded DNA-binding protein', *EMBO Reports*, 3(5), pp. 451-456.

Takeshita, T., Kageyama, S., Furuta, M., Tsuboi, H., Takeuchi, K., Shibata, Y., Shimazaki, Y., Akifusa, S., Ninomiya, T., Kiyohara, Y. and Yamashita, Y. (2016) 'Bacterial diversity in saliva and oral health-related conditions: The Hisayama Study', *Scientific Reports*, 6.

Tang, Y., Schon, E.A., Wilichowski, E., Vazquez-Memije, M.E., Davidson, E. and King, M.P. (2000) 'Rearrangements of Human Mitochondrial DNA (mtDNA): New Insights into the Regulation of mtDNA Copy Number and Gene Expression', *Molecular Biology of the Cell*, 11(4), pp. 1471-1485.

Tanji, K. and Bonilla, E. (2007) 'Optical Imaging Techniques (Histochemical, Immunohistochemical, and In Situ Hybridization Staining Methods) to Visualize Mitochondria', *Methods in Cell Biology*, 80, pp. 135-154.

Taylor, D.R., Zeyl, C. and Cooke, E. (2002) 'Conflicting levels of selection in the accumulation of mitochondrial defects in Saccharomyces cerevisiae', *Proceedings of the National Academy of Sciences of the United States of America*, 99(6), pp. 3690-3694.

Taylor, R.W., Taylor, G.A., Durham, S.E. and Turnbull, D.M. (2001) 'The determination of complete human mitochondrial DNA sequences in single cells: implications for the study of somatic mitochondrial DNA point mutations', *Nucleic Acids Res*, 29(15), pp. E74-4.

Taylor, R.W. and Turnbull, D.M. (2005) 'Mitochondrial DNA mutations in human disease', *Nat Rev Genet*, 6(5), pp. 389-402.

Taylor, S.D., Ericson, N.G., Burton, J.N., Prolla, T.A., Silber, J.R., Shendure, J. and Bielas, J.H. (2014) 'Targeted enrichment and high-resolution digital profiling of mitochondrial DNA deletions in human brain', *Aging Cell*, 13(1), pp. 29-38.

te Velde, E.R., Dorland, M. and Broekmans, F.J. (1998) 'Age at menopause as a marker of reproductive ageing', *Maturitas*, 30(2), pp. 119-125.

Tezze, C., Romanello, V., Desbats, M.A., Fadini, G.P., Albiero, M., Favaro, G., Ciciliot, S., Soriano, M.E., Morbidoni, V., Cerqua, C., Loefler, S., Kern, H., Franceschi, C., Salvioli, S., Conte, M., Blaauw, B., Zampieri, S., Salviati, L., Scorrano, L. and Sandri, M. 'Age-Associated Loss of OPA1 in Muscle Impacts Muscle Mass, Metabolic Homeostasis, Systemic Inflammation, and Epithelial Senescence', *Cell Metabolism*, 25(6), pp. 1374-1389.e6.

Thangaratnarajah, C., Ruprecht, J.J. and Kunji, E.R. (2014) 'Calcium-induced conformational changes of the regulatory domain of human mitochondrial aspartate/glutamate carriers', *Nat Commun*, 5, p. 5491.

The Wellcome Trust Case Control Consortium (2007) 'Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls', *Nature*, 447(7145), pp. 661-678.

Thiede, C., Prange-Krex, G., Freiberg-Richter, J., Bornhauser, M. and Ehninger, G. (2000) 'Buccal swabs but not mouthwash samples can be used to obtain pretransplant DNA fingerprints from recipients of allogeneic bone marrow transplants', *Bone Marrow Transplantation*, 25(5), pp. 575-577.

Thorburn, D.R. (2004) 'Mitochondrial disorders: Prevalence, myths and advances', *Journal of Inherited Metabolic Disease*, 27(3), pp. 349-362.

Thorburn, D.R., Chow, C.W. and Kirby, D.M. (2004) 'Respiratory chain enzyme analysis in muscle and liver', *Mitochondrion*, 4(5–6), pp. 363-375.

Thyagarajan, B., Wang, R., Barcelo, H., Koh, W.-P. and Yuan, J.-M. (2012) 'Mitochondrial Copy Number is Associated with Colorectal Cancer Risk', *Cancer Epidemiology Biomarkers & Prevention*, 21(9), pp. 1574-1581.

Toniolo, P.G., Levitz, M., Zeleniuch-jacquotte, A., Banerjee, S., Koenig, K.L., Shore, R.E., Strax, P. and Pasternack, B.S. (1995) 'A prospective study of endogenous estrogens and breast cancer in postmenopausal women', *Journal of the National Cancer Institute*, 87(3), pp. 190-197.

Torgerson, D.J., Thomas, R.E. and Reid, D.M. (1997) 'Mothers and daughters menopausal ages: is there a link?', *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 74(1), pp. 63-66.

Torroni, A., Bandelt, H.J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M.L., Bonné-Tamir, B. and Scozzari, R. (1998) 'mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe', *American Journal of Human Genetics*, 62(5), pp. 1137-1152.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.L. and Wallace, D.C. (1996) 'Classification of European mtDNAs from an analysis of three European populations', *Genetics*, 144(4), pp. 1835-50.

Torroni, A., Lott, M.T., Cabell, M.F., Chen, Y.S., Lavergne, L. and Wallace, D.C. (1994) 'mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region', *Am J Hum Genet*, 55(4), pp. 760-76.

Torroni, A., Petrozzi, M., D'Urbano, L., Sellitto, D., Zeviani, M., Carrara, F., Carducci, C., Leuzzi, V., Carelli, V., Barboni, P., De Negri, A. and Scozzari, R. (1997) 'Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484', *American Journal of Human Genetics*, 60(5), pp. 1107-1121.

Torroni, A., Richards, M. and Macaulay, V. (2000) 'mtDNA haplogroups and frequency patterns in Europe', *Am J Hum Genet*, 66.

Tranah, G.J., Yokoyama, J.S., Katzman, S.M., Nalls, M.A., Newman, A.B., Harris, T.B., Cesari, M., Manini, T.M., Schork, N.J., Cummings, S.R., Liu, Y., Yaffe, K., Health, A. and Body Composition, S. (2014) 'Mitochondrial DNA sequence associations with dementia and amyloid-beta in elderly African Americans', *Neurobiol Aging*, 35(2), pp. 442 e1-8.

Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J.N., Rovio, A.T., Bruder, C.E., Bohlooly-Y, M., Gidlof, S., Oldfors, A., Wibom, R., Tornell, J., Jacobs, H.T. and

Larsson, N.-G. (2004) 'Premature ageing in mice expressing defective mitochondrial DNA polymerase', *Nature*, 429(6990), pp. 417-423.

Uranova, N., Orlovskaya, D., Vikhreva, O., Zimina, I., Kolomeets, N., Vostrikov, V. and Rachmanova, V. (2001) 'Electron microscopy of oligodendroglia in severe mental illness', *Brain Research Bulletin*, 55(5), pp. 597-610.

Vafai, S.B. and Mootha, V.K. (2012) 'Mitochondrial disorders as windows into an ancient organelle', *Nature*, 491(7424), pp. 374-383.

van Asselt, K.M., Kok, H.S., Putter, H., Wijmenga, C., Peeters, P.H.M., van der Schouw, Y.T., Grobbee, D.E., te Velde, E.R., Mosselman, S. and Pearson, P.L. (2004) 'Linkage Analysis of Extremely Discordant and Concordant Sibling Pairs Identifies Quantitative Trait Loci Influencing Variation in Human Menopausal Age', *The American Journal of Human Genetics*, 74(3), pp. 444-453.

Van Blerkom, J. (2011) 'Mitochondrial function in the human oocyte and embryo and their role in developmental competence', *Mitochondrion*, 11(5), pp. 797-813.

Van den Bogert, C., De Vries, H., Holtrop, M., Muus, P., Dekker, H.L., Van Galen, M.J.M., Bolhuis, P.A. and Taanman, J.W. (1993) 'Regulation of the expression of mitochondrial proteins: relationship between mtDNA copy number and cytochrome-c oxidase activity in human cells and tissues', *BBA - Bioenergetics*, 1144(2), pp. 177-183.

Van Den Eeden, S.K., Tanner, C.M., Bernstein, A.L., Fross, R.D., Leimpeter, A., Bloch, D.A. and Nelson, L.M. (2003) 'Incidence of Parkinson's disease: Variation by age, gender, and race/ethnicity', *American Journal of Epidemiology*, 157(11), pp. 1015-1022.

Van Der Harst, P., Zhang, W., Mateo Leach, I., Rendon, A., Verweij, N., Sehmi, J., Paul, D.S., Elling, U., Allayee, H., Li, X., Radhakrishnan, A., Tan, S.T., Voss, K., Weichenberger, C.X., Albers, C.A., Al-Hussani, A., Asselbergs, F.W., Ciullo, M., Danjou, F., Dina, C., Esko, T., Evans, D.M., Franke, L., Gögele, M., Hartiala, J., Hersch, M., Holm, H., Hottenga, J.J., Kanoni, S., Kleber, M.E., Lagou, V., Langenberg, C., Lopez, L.M., Lyytikäinen, L.P., Melander, O., Murgia, F., Nolte, I.M., O'Reilly, P.F., Padmanabhan, S., Parsa, A., Pirastu, N., Porcu, E., Portas, L., Prokopenko, I., Ried, J.S., Shin, S.Y., Tang, C.S., Teumer, A., Traglia, M., Ulivi, S.,

Westra, H.J., Yang, J., Hua Zhao, J., Anni, F., Abdellaoui, A., Attwood, A., Balkau, B., Bandinelli, S., Bastardot, F., Benyamin, B., Boehm, B.O., Cookson, W.O., Das, D., De Bakker, P.I.W., De Boer, R.A., De Geus, E.J.C., De Moor, M.H., Dimitriou, M., Domingues, F.S., Döring, A., Engström, G., Eyjolfsson, G.I., Ferrucci, L., Fischer, K., Galanello, R., Garner, S.F., Genser, B., Gibson, Q.D., Girotto, G., Gudbjartsson, D.F., Harris, S.E., Hartikainen, A.L., Hastie, C.E., Hedblad, B., Illig, T., Jolley, J., Kähönen, M., Kema, I.P., Kemp, J.P., Liang, L., Lloyd-Jones, H., Loos, R.J.F., Meacham, S., Medland, S.E., Meisinger, C., Memari, Y., Mihailov, E., Miller, K., Moffatt, M.F., Nauck, M., et al. (2012) 'Seventy-five genetic loci influencing the human red blood cell', *Nature*, 492(7429), pp. 369-375.

Van Der Walt, J.M., Dementieva, Y.A., Martin, E.R., Scott, W.K., Nicodemus, K.K., Kroner, C.C., Welsh-Bohmer, K.A., Saunders, A.M., Roses, A.D., Small, G.W., Schmechel, D.E., Murali Doraiswamy, P., Gilbert, J.R., Haines, J.L., Vance, J.M. and Pericak-Vance, M.A. (2004) 'Analysis of European mitochondrial haplogroups with Alzheimer disease risk', *Neuroscience Letters*, 365(1), pp. 28-32.

Van Der Walt, J.M., Nicodemus, K.K., Martin, E.R., Scott, W.K., Nance, M.A., Watts, R.L., Hubble, J.P., Haines, J.L., Koller, W.C., Lyons, K., Pahwa, R., Stern, M.B., Colcher, A., Hiner, B.C., Jankovic, J., Ondo, W.G., Allen Jr, F.H., Goetz, C.G., Small, G.W., Mastaglia, F., Stajich, J.M., McLaurin, A.C., Middleton, L.T., Scott, B.L., Schmechel, D.E., Pericak-Vance, M.A. and Vance, J.M. (2003) 'Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease', *American Journal of Human Genetics*, 72(4), pp. 804-811.

Van Dyck, E., Jank, B., Ragnini, A., Schweyen, R.J., Duyckaerts, C., Sluse, F. and Foury, F. (1995) 'Overexpression of a novel member of the mitochondrial carrier family rescues defects in both DNA and RNA metabolism in yeast mitochondria', *Molecular and General Genetics MGG*, 246(4), pp. 426-436.

Van Goethem, G., Dermaut, B., Löfgren, A., Martin, J.J. and Van Broeckhoven, C. (2001) 'Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions', *Nature Genetics*, 28(3), pp. 211-212.

Van Haute, L., Spits, C., Geens, M., Seneca, S. and Sermon, K. (2013) 'Human embryonic stem cells commonly display large mitochondrial DNA deletions', *Nat Biotech*, 31(1), pp. 20-23.

Van Leuven, J.T., Meister, R.C., Simon, C. and McCutcheon, J.P. (2014) 'Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one', *Cell*, 158(6), pp. 1270-1280.

van Noord, P.A.H., Dubas, J.S., Dorland, M., Boersma, H. and Velde, E.t. (1997) 'Age at natural menopause in a population-based screening cohort: the role of menarche, fecundity, and lifestyle factors', *Fertility and Sterility*, 68(1), pp. 95-102.

van Oven, M. (2015) 'PhyloTree Build 17: Growing the human mitochondrial DNA tree', *Forensic Science International: Genetics Supplement Series*, 5, pp. e392-e394.

van Oven, M. and Kayser, M. (2009) 'Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation', *Human mutation*, 30(2).

Varma, S.D. and Hegde, K.R. (2007) 'Lens thiol depletion by peroxynitrite. Protective effect of pyruvate', *Molecular and Cellular Biochemistry*, 298(1), pp. 199-204.

Vejabhuti Curry, C. (2015) *Medscape*. Available at: http://reference.medscape.com/guide/laboratory-medicine (Accessed: 4th August).

Velarde, M.C. (2014) 'Mitochondrial and sex steroid hormone crosstalk during aging', *Longev Healthspan*, 3.

Veltri, K.L., Espiritu, M. and Singh, G. (1990) 'Distinct genomic copy number in mitochondria of different mammalian organs', *Journal of Cellular Physiology*, 143(1), pp. 160-164.

Venegas, V. and Halberg, M.C. (2012) 'Measurement of Mitochondrial DNA Copy Number', in Wong, P.D.L.-J.C. (ed.) *Mitochondrial Disorders: Biochemical and Molecular Analysis*. Totowa, NJ: Humana Press, pp. 327-335.

Viña, J., Borrás, C., Gambini, J., Sastre, J. and Pallardó, F.V. (2005) 'Why females live longer than males? Importance of the upregulation of longevity-associated genes by oestrogenic compounds', *FEBS Letters*, 579(12), pp. 2541-2545.

Virbasius, J.V. and Scarpulla, R.C. (1994) 'Activation of the human mitochondrial transcription factor A gene by nuclear respiratory factors: A potential regulatory link between nuclear and mitochondrial gene expression in organelle biogenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 91(4), pp. 1309-1313.

Viscomi, C., Bottani, E., Civiletto, G., Cerutti, R., Moggio, M., Fagiolari, G., Schon, Eric A., Lamperti, C. and Zeviani, M. (2011) 'In Vivo Correction of COX Deficiency by Activation of the AMPK/PGC-1α Axis', *Cell Metabolism*, 14(1), pp. 80-90.

Visscher, P.M., Hemani, G., Vinkhuyzen, A.A.E., Chen, G.-B., Lee, S.H., Wray, N.R., Goddard, M.E. and Yang, J. (2014) 'Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples', *PLOS Genetics*, 10(4), p. e1004269.

von der Malsburg, K., Müller, J., Bohnert, M., Oeljeklaus, S., Kwiatkowska, P., Becker, T., Loniewska-Lwowska, A., Wiese, S., Rao, S., Milenkovic, D., Hutu, D., Zerbes, R., Schulze-Specking, A., Meyer, H., Martinou, J.C., Rospert, S., Rehling, P., Meisinger, C., Veenhuis, M., Warscheid, B., van der Klei, I., Pfanner, N., Chacinska, A. and van der Laan, M. (2011) 'Dual Role of Mitofilin in Mitochondrial Membrane Organization and Protein Biogenesis', *Developmental Cell*, 21(4), pp. 694-707.

Vondráčková, A., Veselá, K., Kratochvílová, H., Kučerová Vidrová, V., Vinšová, K., Stránecký, V., Honzík, T., Hansíková, H., Zeman, J. and Tesařová, M. (2014) 'Large copy number variations in combination with point mutations in the TYMP and SCO2 genes found in two patients with mitochondrial disorders', *European Journal of Human Genetics*, 22(3), pp. 431-434.

Vyssokikh, M.Y., Katz, A., Rueck, A., Wuensch, C., DÖRner, A., Zorov, D.B. and Brdiczka, D. (2001) 'Adenine nucleotide translocator isoforms 1 and 2 are differently distributed in the mitochondrial inner membrane and have distinct affinities to cyclophilin D', *Biochemical Journal*, 358(2), pp. 349-358.

Wagenmakers, A.J.M. (1998) 'Muscle Amino Acid Metabolism at Rest and during Exercise: Role in Human Physiology and Metabolism', *Exercise and Sport Sciences Reviews*, 26, pp. 287-314.

Wai, T., Teoli, D. and Shoubridge, E.A. (2008) 'The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes', *Nat Genet*, 40(12), pp. 1484-1488.

Wallace, D.C. (1999) 'Mitochondrial Diseases in Man and Mouse', *Science*, 283(5407), pp. 1482-1488.

Wallace, D.C. 39 (2005a) 'A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A dawn for evolutionary medicine' *Annual Review of Genetics*. pp. 359-407.

Wallace, D.C. (2005b) 'A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine', *Annu Rev Genet*, 39.

Wallace, D.C. (2012) 'Mitochondria and cancer', *Nat Rev Cancer*, 12.

Wallace, D.C. (2013) 'Bioenergetics in human evolution and disease: implications for the origins of biological complexity and the missing genetic variation of common diseases', *Philos Trans R Soc Lond B Biol Sci*, 368(1622), p. 20120267.

Wallace, D.C., Brown, M.D. and Lott, M.T. (1999) 'Mitochondrial DNA variation in human evolution and disease', *Gene*, 238(1), pp. 211-30.

Wallace, D.C. and Chalkia, D. (2013) 'Mitochondrial DNA Genetics and the Heteroplasmy Conundrum in Evolution and Disease', *Cold Spring Harbor Perspectives in Biology*, 5(11), p. a021220.

Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, G.T., Lezza, A.M.S., Elsas Ii, L.J. and Nikoskelainen, E.K. (1988) 'Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy', *Science*, 242(4884), pp. 1427-1430.

Wallace, D.C., Ye, J., Neckelmann, S.N., Singh, G., Webster, K.A. and Greenberg, B.D. (1987) 'Sequence analysis of cDNAs for the human and bovine ATP synthase β subunit: mitochondrial DNA genes sustain seventeen times more mutations', *Current Genetics*, 12(2), pp. 81-90.

Wallace, W.H.B. and Kelsey, T.W. (2010) 'Human Ovarian Reserve from Conception to the Menopause', *PLoS ONE*, 5(1), p. e8772.

Walsh, S.R., Cook, E.J., Goulder, F., Justin, T.A. and Keeling, N.J. (2005) 'Neutrophil-lymphocyte ratio as a prognostic factor in colorectal cancer', *Journal of Surgical Oncology*, 91(3), pp. 181-184.

Wanders, R.J.A., Ruiter, J.P.N., Ijlst, L., Waterham, H.R. and Houten, S.M. (2010) 'The enzymology of mitochondrial fatty acid beta-oxidation and its application to follow-up analysis of positive neonatal screening results', *Journal of Inherited Metabolic Disease*, 33(5), pp. 479-494.

Wang, D., Su, L.Y., Zhang, A.M., Li, Y.Y., Li, X.A., Chen, L.L., Long, H. and Yao, Y.G. (2012) 'Mitochondrial DNA copy number, but not haplogroup, confers a genetic susceptibility to leprosy in Han Chinese from Southwest China', *PLoS ONE*, 7(6).

Wang, K., Li, M. and Hakonarson, H. (2010a) 'Analysing biological pathways in genome-wide association studies', *Nat Rev Genet*, 11(12), pp. 843-854.

Wang, K., Li, M. and Hakonarson, H. (2010b) 'ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data', *Nucleic Acids Res*, 38(16), p. e164.

Wang, Q., Liu, Y. and Zhou, J. (2015) 'Neuroinflammation in Parkinson's disease and its potential as therapeutic target', *Translational Neurodegeneration*, 4, p. 19.

Wang, X. and Gerdes, H.H. (2015) 'Transfer of mitochondria via tunneling nanotubes rescues apoptotic PC12 cells', *Cell Death Differ*, 22(7), pp. 1181-1191.

Wang, X., Zhu, H., Snieder, H., Su, S., Munn, D., Harshfield, G., Maria, B.L., Dong, Y., Treiber, F., Gutin, B. and Shi, H. (2010c) 'Obesity related methylation changes in DNA of peripheral blood leukocytes', *BMC Med*, 8.

Ward, L.D. and Kellis, M. (2012) 'Interpreting noncoding genetic variation in complex traits and human disease', *Nat Biotech*, 30(11), pp. 1095-1106.

Watanabe, A., Arai, M., Koitabashi, N., Niwano, K., Ohyama, Y., Yamada, Y., Kato, N. and Kurabayashi, M. (2011) 'Mitochondrial transcription factors TFAM and TFB2M regulate Serca2 gene transcription', *Cardiovascular Research*, 90(1), pp. 57-67.

Weeber, E.J., Levy, M., Sampson, M.J., Anflous, K., Armstrong, D.L., Brown, S.E., Sweatt, J.D. and Craigen, W.J. (2002) 'The Role of Mitochondrial Porins and the Permeability Transition Pore in Learning and Synaptic Plasticity', *Journal of Biological Chemistry*, 277(21), pp. 18891-18897.

Wei, W., Keogh, M.J., Wilson, I., Coxhead, J., Ryan, S., Rollinson, S., Griffin, H., Kurzawa-Akanbi, M., Santibanez-Koref, M., Talbot, K., Turner, M.R., McKenzie, C.-A., Troakes, C., Attems, J., Smith, C., Al Sarraj, S., Morris, C.M., Ansorge, O., Pickering-Brown, S., Ironside, J.W. and Chinnery, P.F. (2017) 'Mitochondrial DNA point mutations and relative copy number in 1363 disease and control human brains', *Acta Neuropathologica Communications*, 5(1), p. 13.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A. and Schönherr, S. (2016) 'HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing', *Nucleic Acids Research*, 44(W1), pp. W58-W63.

Weitzel, J.M., Iwen, K.A.H. and Seitz, H.J. (2003) 'Regulation of mitochondrial biogenesis by thyroid hormone', *Experimental Physiology*, 88(1), pp. 121-128.

Weizmann Institute of Science (2015) 'GeneCards: The Human Gene Compendium'. Available at: http://www.genecards.org/.

Wellcome Trust Sanger Institute (2008) 'Wellcome Trust Case Control Consortium 2' Wellcome Trust Sanger Institute. April 2008. Wellcome Trust Sanger Institute, Hinxton. Available at: https://www.wtccc.org.uk/ccc2/.

Wellen, K.E., Hatzivassiliou, G., Sachdeva, U.M., Bui, T.V., Cross, J.R. and Thompson, C.B. (2009) 'ATP-citrate lyase links cellular metabolism to histone acetylation', *Science*, 324(5930), pp. 1076-1080.

West, A.P., Khoury-Hanold, W., Staron, M., Tal, M.C., Pineda, C.M., Lang, S.M., Bestwick, M., Duguay, B.A., Raimundo, N., MacDuff, D.A., Kaech, S.M., Smiley, J.R., Means, R.E., Iwasaki, A. and Shadel, G.S. (2015) 'Mitochondrial DNA stress primes the antiviral innate immune response', *Nature*, 520(7548), pp. 553-557.

Westermann, B. (2010) 'Mitochondrial fusion and fission in cell life and death', *Nat Rev Mol Cell Biol*, 11(12), pp. 872-884.

Westermann, B. (2012) 'Bioenergetic role of mitochondrial fusion and fission', *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1817(10), pp. 1833-1838.

Whitacre, C.C. (2001) 'Sex differences in autoimmune disease', *Nat Immunol*, 2(9), pp. 777-780.

Wiedemann, F.R., Manfredi, G., Mawrin, C., Beal, M.F. and Schon, E.A. (2002) 'Mitochondrial DNA and respiratory chain function in spinal cords of ALS patients', *J Neurochem*, 80(4), pp. 616-25.

Wilder, R.L. (1995) 'Neuroendocrine-Immune System Interactions and Autoimmunity', *Annual Review of Immunology*, 13(1), pp. 307-338.

Williams, R.S. (1986) 'Mitochondrial gene expression in mammalian striated muscle. Evidence that variation in gene dosage is the major regulatory event', *Journal of Biological Chemistry*, 261(26), pp. 12390-12394.

Winklhofer, K.F. and Haass, C. (2010) 'Mitochondrial dysfunction in Parkinson's disease', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1802(1), pp. 29-44.

Wittig, I., Braun, H.-P. and Schagger, H. (2006) 'Blue native PAGE', *Nat. Protocols*, 1(1), pp. 418-428.

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J.a., Kutalik, Z., Amin, N., Buchkovich, M.L., Croteau-Chonka, D.C., Day, F.R., Duan, Y., Fall, T., Fehrmann, R., Ferreira, T., Jackson, A.U., Karjalainen, J., Lo, K.S., Locke, A.E., Magi, R., Mihailov, E., Porcu, E., Randall, J.C., Scherag, A., Vinkhuyzen, A.A.E., Westra, H.-J., Winkler, T.W., Workalemahu, T., Zhao, J.H., Absher, D., Albrecht, E., Anderson, D., Baron, J., Beekman, M., Demirkan, A., Ehret, G.B., Feenstra, B., Feitosa, M.F., Fischer, K., Fraser, R.M., Goel, A., Gong, J., Justice, A.E., Kanoni, S., Kleber, M.E., Kristiansson, K., Lim, U., Lotay, V., Lui, J.C., Mangino, M., Leach, I.M., Medina-Gomez, C., Nalls, M.A., Nyholt, D.R., Palmer, C.D., Pasko, D., Pechlivanis, S., Prokopenko, I., Ried, J.S., Ripke, S., Shungin, D., Stancakova, A., Strawbridge, R.J., Sung, Y.J., Tanaka, T., Teumer, A., Trompet, S., van der Laan, S.W., van Setten, J., Van Vliet-Ostaptchouk, J.V., Wang, Z., Yengo, L., Zhang, W., Afzal, U., Arnlov, J., Arscott, G.M., Bandinelli, S., Barrett, A., Bellis, C., Bennett, A.J., Berne, C., Bluher, M., Bolton, J.L., Bottcher, Y., Boyd, H.A., Bruinenberg, M., Buckley, B.M., Buyske, S., Caspersen, I.H., Chines, P.S., Clarke, R., Claudi-Boehm, S., Cooper, M., Daw, E.W., De Jong, P.A., Deelen, J., Delgado, G., et al. (2014) 'Defining the role of common variation in the genomic and biological architecture of adult human height', *Nat Genet*, 46(11), pp. 1173-1186.

Woods, A., Dickerson, K., Heath, R., Hong, S.P., Momcilovic, M., Johnstone, S.R., Carlson, M. and Carling, D. (2005) 'Ca2+/calmodulin-dependent protein kinase kinase-β acts upstream of AMP-activated protein kinase in mammalian cells', *Cell Metabolism*, 2(1), pp. 21-33.

Wu, B. and Pankow, J.S. (2015) 'Statistical methods for association tests of multiple continuous traits in genome-wide association studies', *Annals of human genetics*, 79(4), pp. 282-293.

Wu, G. (2009) 'Amino acids: metabolism, functions, and nutrition', *Amino Acids*, 37(1), pp. 1-17.

Wu, Z., Puigserver, P., Andersson, U., Zhang, C., Adelmant, G., Mootha, V., Troy, A., Cinti, S., Lowell, B., Scarpulla, R.C. and Spiegelman, B.M. (1999) 'Mechanisms Controlling Mitochondrial Biogenesis and Respiration through the Thermogenic Coactivator PGC-1', *Cell*, 98(1), pp. 115-124.

Xia, P., An, H.-X., Dang, C.-X., Radpour, R., Kohler, C., Fokas, E., Engenhart-Cabillic, R., Holzgreve, W. and Zhong, X.Y. (2009) 'Decreased mitochondrial DNA content in blood samples of patients with stage I breast cancer', *BMC Cancer*, 9(1), p. 454.

Xing, J., Chen, M., Wood, C.G., Lin, J., Spitz, M.R., Ma, J., Amos, C.I., Shields, P.G., Benowitz, N.L., Gu, J., de Andrade, M., Swan, G.E. and Wu, X. (2008) 'Mitochondrial DNA Content: Its Genetic Heritability and Association With Renal Cell Carcinoma', *JNCI Journal of the National Cancer Institute*, 100(15), pp. 1104-1112.

Yang, L., Tan, Z., Wang, D., Xue, L., Guan, M.-x., Huang, T. and Li, R. (2014) 'Species identification through mitochondrial rRNA genetic analysis', *Scientific Reports*, 4, p. 4089.

Yang, R.-F., Sun, L.-H., Zhang, R., Zhang, Y., Luo, Y.-X., Zheng, W., Zhang, Z.-Q., Chen, H.-Z. and Liu, D.-P. (2015) 'Suppression of Mic60 compromises mitochondrial transcription and oxidative phosphorylation', *Sci. Rep.*, 5.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S.J., Cunningham, F., Aken, B.L., Zerbino, D.R.

and Flicek, P. (2016) 'Ensembl 2016', *Nucleic Acids Research*, 44(D1), pp. D710-D716.

Ye, W., Tang, X., Liu, C., Wen, C., Li, W. and Lyu, J. (2017) 'Accurate quantitation of circulating cell-free mitochondrial DNA in plasma by droplet digital PCR', *Analytical and Bioanalytical Chemistry*, 409(10), pp. 2727-2735.

Yin, P.H., Lee, H.C., Chau, G.Y., Wu, Y.T., Li, S.H., Lui, W.Y., Wei, Y.H., Liu, T.Y. and Chi, C.W. (2004) 'Alteration of the copy number and deletion of mitochondrial DNA in human hepatocellular carcinoma', *British Journal of Cancer*, 90(12), pp. 2390-2396.

Ylikallio, E., Tyynismaa, H., Tsutsui, H., Ide, T. and Suomalainen, A. (2010) 'High mitochondrial DNA copy number has detrimental effects in mice', *Human Molecular Genetics*, 19(13), pp. 2695-2705.

Youle, R.J. and Narendra, D.P. (2011) 'Mechanisms of mitophagy', *Nature Reviews Molecular Cell Biology*, 12(1), pp. 9-14.

Young, J.M. and McNeilly, A.S. (2010) 'Theca: the forgotten cell of the ovarian follicle', *Reproduction*, 140(4), pp. 489-504.

Yu-Wai-Man, P., Griffiths, P.G. and Chinnery, P.F. (2011) 'Mitochondrial optic neuropathies - Disease mechanisms and therapeutic strategies', *Progress in Retinal and Eye Research*, 30(2), pp. 81-114.

Yu, J.-W. and Lee, M.-S. (2016) 'Mitochondria and the NLRP3 inflammasome: physiological and pathological relevance', *Archives of Pharmacal Research*, 39(11), pp. 1503-1518.

Yu, M. (2012) 'Circulating cell-free mitochondrial DNA as a novel cancer biomarker: opportunities and challenges', *Mitochondrial DNA*, 23(5), pp. 329-332.

Zerial, M. and McBride, H. (2001) 'Rab proteins as membrane organizers', *Nature reviews. Molecular cell biology*, 2(2), pp. 107-117.

Zeviani, M. and Antozzi, C. (1997) 'Mitochondrial disorders', *Mol Hum Reprod*, 3.

Zeviani, M. and Di Donato, S. (2004) 'Mitochondrial disorders', *Brain*, 127(10), pp. 2153-2172.

Zhang, F., Qi, Y., Zhou, K., Zhang, G., Linask, K. and Xu, H. (2015) 'The cAMP phosphodiesterase Prune localizes to the mitochondrial matrix and promotes mtDNA replication by stabilizing TFAM', *EMBO Reports*, 16(4), pp. 520-527.

Zhang, Y., Bharathi, S.S., Rardin, M.J., Lu, J., Maringer, K.V., Sims-Lucas, S., Prochownik, E.V., Gibson, B.W. and Goetzman, E.S. (2017) 'SIRT5 Binds to Cardiolipin and Regulates the Electron Transport Chain', *Journal of Biological Chemistry*.

Zheng, B., Liao, Z., Locascio, J.J., Lesniak, K.A., Roderick, S.S., Watt, M.L., Eklund, A.C., Zhang-James, Y., Kim, P.D., Hauser, M.A., Grünblatt, E., Moran, L.B., Mandel, S.A., Riederer, P., Miller, R.M., Federoff, H.J., Wüllner, U., Papapetropoulos, S., Youdim, M.B., Cantuti-Castelvetri, I., Young, A.B., Vance, J.M., Davis, R.L., Hedreen, J.C., Adler, C.H., Beach, T.G., Graeber, M.B., Middleton, F.A., Rochet, J.-C., Scherzer, C.R. and the Global, P.D.G.E.C. (2010) 'PGC-1α, A Potential Therapeutic Target for Early Intervention in Parkinson's Disease', *Science translational medicine*, 2(52), pp. 52ra73-52ra73.

Zheng, J., Li, Y., Abecasis, G.R. and Scheet, P. (2011) 'A Comparison of Approaches to Account for Uncertainty in Analysis of Imputed Genotypes', *Genetic epidemiology*, 35(2), pp. 102-110.

Zhou, R., Yazdi, A.S., Menu, P. and Tschopp, J. (2011) 'A role for mitochondria in NLRP3 inflammasome activation', *Nature*, 469(7329), pp. 221-225.

Zieliński, Ł.P., Smith, A.C., Smith, A.G. and Robinson, A.J. (2016) 'Metabolic flexibility of mitochondrial respiratory chain disorders predicted by computer modelling', *Mitochondrion*, 31, pp. 45-55.

Zole, E., Zadinane, K., Pliss, L. and Ranka, R. (2017) 'Linkage between mitochondrial genome alterations, telomere length and aging population', *Mitochondrial DNA Part A*, pp. 1-8.

Zuniga, R.M., Torcuator, R., Jain, R., Anderson, J., Doyle, T., Schultz, L. and Mikkelsen, T. (2010) 'Rebound tumour progression after the cessation of bevacizumab therapy in patients with recurrent high-grade glioma', *J Neurooncol*, 99(2), pp. 237-42.