Tampere University

JAAKKO LAAKSONEN

Mitochondrial Genetic Determinants of Peripheral Blood Transcriptomics, DNA Methylation and Blood Pressure

Tampere University Dissertations 665

Tampere University Dissertations 665

JAAKKO LAAKSONEN

Mitochondrial Genetic Determinants of Peripheral Blood Transcriptomics, DNA Methylation and Blood Pressure

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Medicine and Health Technology
of Tampere University,
for public discussion in the auditorium F114
of the Arvo Building, Arvo Ylpön katu 34, Tampere,
on 18 November 2022, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Medicine and Health Technology
Finland

Responsible supervisor and Custos	Professor Terho Lehtimäki Tampere University Finland	
Pre-examiners	Associate Professor Reetta Hinttala University of Oulu Finland	Adjunct Professor Jaakko Pohjoismäki University of Eastern Finland Finland
Opponent	Academy Professor Anu Wartiovaara University of Helsinki Finland	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Copyright ©2022 Jaakko Laaksonen

Cover design: Roihu Inc.

ISBN 978-952-03-2554-1 (print)
ISBN 978-952-03-2555-8 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-2555-8

Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino
Joensuu 2022

ACKNOWLEDGEMENTS

The studies presented in this thesis were carried out at the Faculty of Medicine and Health Technology, Tampere University, at the Department of Clinical Chemistry during the years 2016–2021. Many great people have helped me along the way.

First of all, I want to thank my supervisor, Professor Terho Lehtimäki. His enthusiasm for science has been contagious, and this work would never have been possible without his continuous encouragement and excellent connections to other research groups. I am also grateful for the freedom he has given me to focus on topics that I have found most interesting.

The official reviewers of the thesis manuscript, Associate Professor Reetta Hinttala and Senior Researcher Jaakko Pohjoismäki, are warmly thanked for their valuable comments that improved the thesis.

Professor Mika Kähönen, co-author and member of my thesis committee, and Professor Jussi Hernesniemi, member of my thesis committee, are also gratefully acknowledged.

I wish to thank Ilkka Seppälä for teaching me the essential skills in R programming and Pashupati Mishra for his valuable support in the bioinformatic analyses. Emma Raitoharju and Nina Mononen deserve my sincere thanks for all the help I have received over the years. Other co-authors in our group, Leo-Pekka Lyytikäinen and Saara Marttila, are also warmly thanked.

This work could not have been possible without those who have been involved in the YFS, FINCAVAS and LURIC studies. In particular, I would like to thank Academy Professor Olli Raitakari, Professor Markus Juonala, Nina Hutri-Kähönen, Marcus Kleber and Professor Winfried März for the valuable work they have done with the cohorts. I also warmly thank all the other co-authors not mentioned earlier: Maija Lepistö, Henrikki Almusa, Pekka Ellonen, Professor Emeritus Jukka T. Salonen, Melanie Waldenberger, Professor Thomas Illig, Graciela Delgado and Professor Stefan Lorkowski. In addition, I want to thank all the cohort participants in

the aforementioned studies.

My parents Leena and Risto and my brother Joonas deserve my gratitude for their unwavering support and encouragement throughout my studies. My grandparents Unto, who unfortunately passed away while I was finishing this thesis, and Maija-Liisa also have always cheered me on, for which I am grateful. I would also like to thank my parents-in-law, Marjo and Erkka, for their support. An honourable mention is given to all of my friends that I have had around me through this project and life in general: you are awesome, thank you.

The most important ones are saved for last. My deepest thankfulness goes to my best friend and life companion Marleena for her love and peer support. We have now both completed our separate doctoral thesis journeys, but together we have managed to create something much more precious: our daughter Armi, who brings sheer, unimaginable joy to my life.

This thesis project was supported by personal grants awarded by the Juhani Aho Foundation, Laboratoriolääketieteen edistämissäätiö, the Aarne Koskelo Foundation, and the Finnish Medical Foundation.

Jyväskylä, August 2022

Jaakko Laaksonen

ABSTRACT

Human mitochondrial DNA (mtDNA) is a ~16.5 kb circular molecule encoding 37 genes, including 13 protein subunits of the oxidative phosphorylation complexes responsible for producing the majority of adenosine triphosphate that cells use for chemical energy. The vast majority of the mitochondria-destined proteins, however, are encoded by nuclear DNA, and in order to adapt to the ever-changing cellular milieu, these two genetic compartments maintain a bi-directional regulation, leading to alterations in DNA methylation (DNAm) and gene expression. Genetic variation in the mtDNA is also capable of affecting DNAm and gene expression, but the effects in natural populations are not completely known. As mitochondria lie at an interface between the bioenergetic processes of the human body, they are potent mediators of common metabolic disorders. The maternal inheritance of mtDNA has also created a hypothesis that mtDNA variation may have sex-specific effects on specific traits.

The aims of this study were to examine the associations of mtDNA single-nucleotide polymorphisms (mtSNPs) with genome-wide peripheral blood transcriptome and DNA methylation profiles and to investigate whether these associations show sexual dimorphism or are affected by the onset of prediabetes, a condition preceding type 2 diabetes mellitus. The mitochondrial genetic determinants of blood pressure were also examined.

In study I, genome-wide peripheral blood transcript data and mtSNPs obtained by next-generation sequencing from a population-based Young Finns Study (YFS) cohort ($n = 955$) were used. In study II, a discovery association study on nuclear DNAm was performed in the YFS population ($n = 926$), and replication was sought in the Ludwigshafen Risk and Cardiovascular Health (LURIC) study ($n = 2,317$). In study III, mitochondrial genetic associations with blood pressure were studied in the YFS ($n = 1,150$) and in the Finnish Cardiovascular Study (FINCAVAS; $n = 3,112$) by a meta-analysis. In the LURIC and FINCAVAS populations, mtSNPs were ob-

tained with microarrays. The sexual dimorphism was examined by applying the analyses separately to males and females and testing for significant differences in the effect size (I–III). The prediabetes-specific effects were investigated similarly by performing the analyses separately to individuals with prediabetes and normoglycaemia (I–II).

Study I identified 53 associations between mtSNPs and gene transcripts, corresponding to 7 genes and 31 variants. Some associations were also replications of previously reported results. Eight associations remained significant after conditional analysis. In addition, five genes showed differential expression between haplogroups. One association demonstrated prediabetes-specific effects, while no evidence for sex-specific effects on gene expression was observed. In study II, numerous methylation quantitative trait mtDNA loci were observed in the discovery phase. Of these, 19 variant-based and four haplogroup-based associations were replicated and reached epigenome-wide significance when the results were combined in a meta-analysis. Two differentially methylated sites associated with gene transcripts in the YFS. Discovery analysis also showed that several associations had sex- or prediabetes-specific effects, but none of them were replicated in the LURIC population. Study III did not identify any variants associated with blood pressure levels nor any evidence of sexual dimorphism.

These results both reveal new and replicate previously reported genetic associations with peripheral blood transcriptomics. For the first time on a cohort level, mitochondrial genetic determinants of DNAm are presented. Some of the transcripts and CpG sites may be linked to the biological processes taking place in mitochondria, which suggests that the associations represent the mitochondrial–nuclear communication in order to maintain cellular homeostasis. The variant associations with transcriptomics and DNAm did not correspond to each other, which indicates that, if there is a causal relationship between mtDNA variation and peripheral blood transcriptomics, the regulatory mechanisms are not mediated by changes in DNAm. As suggested by several previous studies, mtDNA variation does not seem to have a significant role in the regulation of blood pressure. Finally, no convincing evidence was found of sex- and prediabetes-specific effects of mtDNA variation.

TIIVISTELMÄ

Ihmisen mitokondrio-DNA (mtDNA) on rengasmainen, n. 16 500 emäsparin mittainen 37:ää eri geeniä koodaava molekyyli. Kolmetoista näistä geeneistä muodostaa osia oksidatiivisen fosforylaation entsyymikomplekseihin, joiden avulla soluille tuotetaan kemiallista energiaa adenosiinitrifosfaatin muodossa. Suurin osa mitokondriossa vaikuttavista proteiineista on kuitenkin tuman DNA:n koodaamia. Jotta solujen toiminta sopeutuu asianmukaisesti alati muuttuviin olosuhteisiin, näiden kahden genomin välinen kahdensuuntainen vuorovaikutus saa aikaan muutoksia DNA:n metylaatiossa ja geenien ilmentymisessä. Myös mtDNA:n geneettisellä muuntelulla on osoitettu olevan oma vaikutuksensa näihin muutoksiin, mutta väestötasolla vaikutukset eivät kuitenkaan ole täysin tunnettuja. Mitokondrioiden keskeinen rooli elimistön aineenvaihdunnassa korostaa niiden mahdollista myötävaikutusta metabolisten häiriöiden kehittymisessä. MtDNA:n maternaalinen periytyminen on myös johtanut hypoteesiin, jonka mukaan mtDNA:n muuntelulla saattaa olla sukupuolen mukaan määräytyviä vaikutuksia.

Tämän työn tavoitteena oli tarkastella mtDNA:n yhden emäksen polymorfismien (mtSNP) assosiaatioita kokoverestä mitattuun genomin laajuiseen geenien ilmentyiseen ja DNA:n metylaatioon. Lisäksi tavoitteena oli selvittää, vaikuttaako sukupuoli tai prediabetes, tyypin 2 diabeteksen esiaste, osaltaan näihin assosiaatioihin. Lisäksi tutkittiin mtDNA:n muuntelun yhteyttä verenpaineeseen.

Osatyössä I käytettiin Lasten sepelvaltimotaudin riskitekijät (LASERI) -aineistosta ($n = 955$) mitattua transkriptiodataa sekä rinnakkaissekvensoinnin avulla määritettyjä mtDNA-variantteja. Osatyössä II mtSNP:iden ja DNA:n metylaatiotasojen välisiä assosiaatioita selvitettiin ensin LASERI-aineistossa ($n = 926$), minkä jälkeen löydöksiä pyrittiin toistamaan Ludwigshafen Risk and Cardiovascular Health (LURIC) -kohortissa ($n = 2317$). Osatyössä III tutkittiin mtSNP:iden yhteyttä verenpaineeseen LASERI- ($n = 1150$) ja Finnish Cardiovascular Study (FINCAVAS) -aineistoissa ($n = 3112$) meta-analyysin keinoin. MtSNP:t määritettiin LURIC- ja FINCAVAS-

aineistoissa DNA-mikrosiruja käyttämällä. Sukupuolen (I–III) ja prediabeteksen (I ja II) vaikutusta tutkittiin analysoimalla erikseen miehet ja naiset sekä prediabeetikot ja normoglykeemiset henkilöt ja vertailemalla assosiaatioiden efektikokoja ryhmien välillä.

Osatyössä I havaitiin yhteensä 53 merkitsevää assosiaatiota mtSNP:iden ja transkriptien välillä, koostuen seitsemästä eri geenistä ja 31 variantista. Eräät assosiaatiot oli jo aiemmin havaittu muissa tutkimuksissa, ja tässä työssä ne kyettiin replikoimaan. Ehdollistamisanalyysin jälkeen kahdeksan löydöstä säilyi merkitsevänä. Lisäksi viiden geenin ekpressiotasoissa oli eroja haploryhmien välillä. Prediabetes vaikutti yhden mtSNP:n ja geenitranskriptin väliseen yhteyteen, mutta sukupuolella ei ollut merkitsevää vaikutusta. Osatyössä II LASERI-aineistossa havaittiin aluksi lukuisia assosiaatiosignaaleja mtDNA-variaation ja metylaation välillä, joista 19 mtSNP- ja neljä haploryhmäsignaalia replikoitui LURIC-aineistossa ja säilyi lisäksi meta-analyysissä merkitsevänä koko epigenomin tasolla. Kaksi tunnistetuista metylaatiokohdista oli yhteydessä geeniekspressioon LASERI-aineistossa. Samassa aineistossa havaittiin myös useita assosiaatioita, joihin sukupuolella tai prediabeteksellä oli merkitsevä vaikutus, mutta nämä tulokset eivät toistuneet LURIC-aineistossa. Osatyössä III ei havaitu viitteitä siitä, että mtDNA:n variaatiolla olisi merkitsevää yhteyttä verenpaineeseen tai että sukupuolikaan vaikuttaisi näihin assosiaatiosignaaleihin.

Tässä työssä sekä osoitetaan uusia että replikoidaan jo aiemmissa tutkimuksissa julkaistuja mtDNA:n variaation ja geeniekspression välisiä assosiaatiosignaaleja, ja DNA:n metylaation ja mtDNA-variaation välisiä yhteyksiä tutkitaan ensimmäistä kertaa väestötason kohorteissa. Osa tunnistetuista geenitranskripteistä ja metylaatiokohdista voidaan yhdistää mitokondrioissa tapahtuviin prosesseihin, ja tulokset voivatkin edustaa tuman ja mitokondrioiden välistä, solutason homeostaasin ylläpitämiseksi tapahtuvaa signalointia. Assosiaatiosignaalit geeniekspression ja metylaation välillä eivät kuitenkaan vastanneet toisiaan. Tämä tarkoittaa, että mikäli mtDNA-variaation ja geeniekspression välillä on kausaalinen yhteys, säätelymekanismit eivät ole metylaatiovälitteisiä. Useat aiemmatkin tutkimukset ovat viitanneet siihen, ettei mtDNA:n variaatiolla näytä olevan merkittävää vaikutusta verenpaineeseen. Vakuuttavaa osoitusta sukupuolen tai prediabeteksen vaikutuksesta edellä mainittuihin assosiaatiosignaaleihin ei havaittu.

CONTENTS

1 Introduction 19

2 Review of the literature 21
2.1 Mitochondrial form and function 21
2.2 Mitochondrial DNA (mtDNA) 23
2.2.1 Variation of mtDNA 26
2.2.2 Common variants and mitochondrial haplogroups 29
2.3 Gene expression and epigenetics 30
2.3.1 DNA transcription 30
2.3.2 DNA methylation 31
2.4 Mitochondrial control of gene expression and DNA methylation . . 32
2.4.1 Mechanisms of mitochondrial–nuclear communication 33
2.4.2 Role of mtDNA variants 37
2.4.3 Sexual dimorphism 40
2.5 Cardiometabolic risk factors and mtDNA variants 41
2.5.1 Diabetes-related changes in gene expression and DNA methylation 41
2.5.2 Hypertension-associated variants 42

3 Aims of the study 45

4 Materials and methods 47
4.1 Study cohorts 47
4.1.1 The Young Finns Study (I–III) 47
4.1.2 The Ludwigshafen Risk and Cardiovascular Health study (II) 47
4.1.3 The Finnish Cardiovascular Study (III) 48
4.1.4 Ethical considerations (I–III) 48

4.2 Identification of mtDNA variants 48
4.2.1 Next-generation sequencing (I–III) 48
4.2.2 Genotyping arrays (II and III) 49
4.2.3 Haplogroup classification (I and II) 50
4.3 Genome-wide transcriptome analysis (I and II) 50
4.4 Epigenome-wide DNA methylation analysis (II) 51
4.5 Definition of clinical phenotypes 51
4.5.1 Prediabetes (I and II) 51
4.5.2 Smoking (II) 52
4.5.3 Blood pressure (III) 52
4.6 Statistical methods 53
4.6.1 Association analyses on the genome-wide transcriptome (I) . 53
4.6.2 Association analyses on DNA methylation (II) 54
4.6.3 Association analyses on blood pressure (III) 57

5 Results 59
5.1 Associations of mtDNA variants with the genome-wide transcriptome (I) 59
5.2 Associations of mtDNA variants with DNA methylation (II) 63
5.3 Prediabetes-specific effects (I and II) 67
5.4 Mitochondrial GWAS on blood pressure (III) 68
5.5 Sex-specific effects (I–III) 69

6 Discussion 71
6.1 Transcriptomic changes associated with mtDNA variation (I) 71
6.2 DNA methylation changes associated with mtDNA variation (II) . . 72
6.3 Prediabetes-specific effects (I and II) 75
6.4 Role of mtDNA variation in blood pressure (III) 75
6.5 No evidence of sexually dimorphic associations (I–III) 76
6.6 Strengths and limitations 77
6.7 Future perspectives 79

7 Conclusions 81

References 83

Publication I 115
Publication II 129
Publication III 145

ABBREVIATIONS

α-KG	α-ketoglutarate
λ_B	Bayesian inflation factor
λ_G	genomic inflation factor
2-HG	2-hydroxyglutarate
5mC	5-methylcytosine
acetyl-CoA	acetyl coenzyme A
ADP	adenosine diphosphate
AMPK	adenosine monophosphate-activated protein kinase
ATP	adenosine triphosphate
BMI	body mass index
BP	blood pressure
CpG	cytosine-phosphate-guanine, sequence of CG in a nucleotide chain
DBP	diastolic blood pressure
DNA	deoxyribonucleic acid
DNAm	DNA methylation
DNMT	DNA methyltransferase
ETC	electron transport chain
EWAS	epigenome-wide association study
FAD/FADH2	flavin adenine dinucleotide/its reduced form
FINCAVAS	Finnish Cardiovascular Study

FPG	fasting plasma glucose
GWAS	genome-wide association study
GWE	genome-wide expression
H strand	heavy strand
HbA1c	glycated hemoglobin
HV	hypervariable
JMJD	Jumonji C domain-containing protein
JNK	c-Jun N-terminal kinase
L strand	light strand
LD	linkage disequilibrium
LHON	Leber's hereditary optic neuropathy
LURIC	Ludwigshafen Risk and Cardiovascular Health
MAP	mean arterial pressure
MDP	mitochondrial-derived peptide
MNX	mitochondrial–nuclear exchange
mRNA	messenger ribonucleic acid
mtDNA	mitochondrial deoxyribonucleic acid
mtSNP	mitochondrial single nucleotide polymorphism
NAD/NADH	nicotinamide adenine dinucleotide/its reduced form
nDNA	nuclear deoxyribonucleic acid
NF-kB	nuclear factor kappa-B
NFATC	nuclear factor of activated T cells
NFE2L2	nuclear factor erythroid 2-related factor 2
NGS	next-generation sequencing
NuMTs	nuclear–mitochondrial sequences
OGTT	oral glucose tolerance test
OXPHOS	oxidative phosphorylation

PC	principal component
PGC1α	peroxisome proliferator-activated receptor gamma coactivator 1-alpha
QC	quality control
QTL	quantitative trait loci
rCRS	revised Cambridge Reference Sequence
ROS	reactive oxygen species
rRNA	ribosomal ribonucleic acid
RSRS	Reconstructed Sapiens Reference Sequence
SAM	S-adenosyl methionine
SBP	systolic blood pressure
SE	standard error
SKAT	sequence kernel association test
T2DM	type 2 diabetes mellitus
TCA	tricarboxylic acid
TET	ten-eleven translocation
tRNA	transfer ribonucleic acid
VAF	variant allele frequency
YFS	Young Finns Study

ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, referred to in the text with their Roman numerals I–III.

Publication I — Laaksonen J, Seppälä I, Raitoharju E, Mononen N, Lyytikäinen LP, Waldenberger M, Illig T, Lepistö M, Almusa H, Ellonen P, Hutri-Kähönen N, Juonala M, Kähönen M, Raitakari O, Salonen JT, Lehtimäki T (2019): Discovery of mitochondrial DNA variants associated with genome-wide blood cell gene expression: a population-based mtDNA sequencing study. *Human Molecular Genetics* 28(8):1381–1391

Publication II — Laaksonen J, Mishra PP, Seppälä I, Raitoharju E, Marttila S, Mononen N, Lyytikäinen LP, Kleber ME, Delgado GE, Lepistö M, Almusa H, Ellonen P, Lorkowski S, März W, Hutri-Kähönen N, Raitakari O, Kähönen M, Salonen JT, Lehtimäki T (2022): Mitochondrial genome-wide analysis of nuclear DNA methylation quantitative trait loci. *Human Molecular Genetics* 31(10):1720–1732

Publication III — Laaksonen J, Mishra PP, Seppälä I, Lyytikäinen LP, Raitoharju E, Mononen N, Lepistö M, Almusa H, Ellonen P, Hutri-Kähönen N, Juonala M, Raitakari O, Kähönen M, Salonen JT, Lehtimäki T (2021): Examining the effect of mitochondrial DNA variants on blood pressure in two Finnish cohorts. *Scientific Reports* 11:611

Author's contribution

In all three original studies, the author was the main and corresponding author, contributing to the planning of the study, designing and executing the statistical analyses, interpreting the results, and writing and editing the original manuscripts. In studies II and III, the author also analysed the previously collected mitochondrial DNA sequencing data.

1 INTRODUCTION

Mitochondria – from the Greek *mitos* (thread-like) and *khondros* (grain or granule) – are cellular organelles with a myriad of functions. To reiterate an often-used cliché, they are 'the powerhouses of the cell', as they produce the majority of a cell's adenosine triphosphate (ATP) through oxidative phosphorylation (OXPHOS). (Schon et al. 2012.) Beyond cellular energy production, mitochondria mediate several key cellular functions, such as calcium signalling, the regulation of cellular metabolism via haem and steroid synthesis, and programmed cell death. Many fundamental metabolic pathways, such as the tricarboxylic acid (TCA) cycle and the fatty acid β-oxidation, take place in mitochondria. (Giacomello et al. 2020; Pfanner et al. 2019.)

Human mitochondrial DNA (mtDNA) is a maternally inherited, double-stranded circular molecule containing 37 genes with high sequence variability. While there are many examples of pathogenic mtDNA polymorphisms resulting in severe disease, most variants are benign, only slightly pathogenic or even beneficial, resulting in changes only at a cellular level. (Stewart and Chinnery 2021.)

Bidirectional communication exists between nuclear and mitochondrial genomes in order to maintain optimal cellular function under different physiological and pathological conditions (Mottis et al. 2019). MtDNA variants have been observed to associate with nuclear gene expression in studies *in vitro* (Kenney et al. 2014b), in common mouse models (Dunham-Snary et al. 2018), and in mitochondrial genome-wide association studies (GWASs) (Kassam et al. 2016). The hypothesis that mtDNA variation can modulate the nuclear DNA (nDNA) methylation profiles has been strengthened by studies *in vitro* (Atilano et al. 2015; Kopinski et al. 2019) and a small study of articular cartilage (Cortés-Pereira et al. 2019), but so far no evidence exist from larger-scale GWASs.

The maternal inheritance of mtDNA has raised a hypothesis that mtDNA mutations may accumulate within generations if they show male-specific effects. A study of *Drosophila melanogaster* found evidence of sex-specific mitochondrial genetic con-

trol of gene expression (Innocenti et al. 2011), but in humans, evidence is lacking (Kassam et al. 2016).

Under specific physiological, environmental and other genetic circumstances, mtDNA variants may modulate the susceptibility to common multifactorial diseases. Definitive answers regarding the role of mtDNA variation in blood pressure levels are lacking, as both positive and negative results have been published (Buford et al. 2018; Saxena et al. 2006). MtDNA variants have been reported to contribute to glycaemic traits, possibly via affecting the expression of nuclear genes regulating cellular energetics (Hwang et al. 2011; Kraja et al. 2019).

This work aims to: **a)** examine the associations of mtDNA variants with genome-wide peripheral blood transcriptome and DNA methylation profiles by mitochondrial GWASs; **b)** investigate whether these associations show sexual dimorphism or are affected by the onset of prediabetes, a condition preceding type 2 diabetes mellitus (T2DM); and **c)** study the mitochondrial genetic determinants of blood pressure and their sex-specific effects.

2 REVIEW OF THE LITERATURE

2.1 Mitochondrial form and function

Mitochondria are surrounded by a double-membrane structure, consisting of inner and outer mitochondrial membranes separated by an intermembrane space. The inner membrane forms several folds (cristae), extending into the mitochondrial matrix and increasing the surface area. The matrix contains the mitochondrial genome, ribosomes, transfer ribonucleic acid (tRNA) and enzymes taking part in the reactions of oxidative metabolism. The outer membrane has many protein-based pores, allowing the passage of ions and molecules. (Friedman and Nunnari 2014.) Though often depicted as discrete bean-like structures, as in Figure 2.1, in reality mitochondria form a dynamic network in which they undergo continual cycles of fusion and fission (Giacomello et al. 2020).

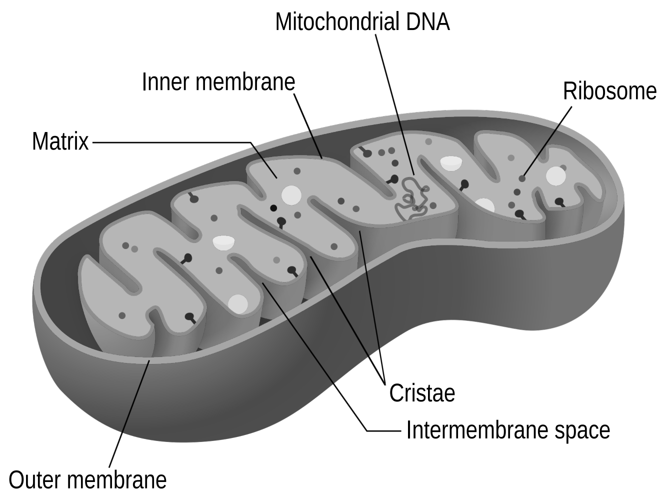

Figure 2.1 Schematic structure of a mitochondrion. Derived from 'Mitochondrion mini' by Kelvin Ma, (CC0 1.0), via Wikimedia Commons.

The core process in OXPHOS is the electron transport chain (ETC) which consists of four protein complexes (I–IV) integrated into the inner mitochondrial membrane. The TCA cycle produces reduced nicotinamide adenine dinucleotide (NADH) and reduced flavin adenine dinucleotide (FADH2), which donate electrons to the ETC via OXPHOS complexes I and II, respectively. At complex I, the transfer of electrons results in the pumping of protons (H^+ ions) from the matrix into the intermembrane space. Complex II does not span the inner membrane, and no protons are translocated by it. (Nolfi-Donegan et al. 2020; Zhao et al. 2019.) An overview of the OXPHOS system in shown in Figure 2.2.

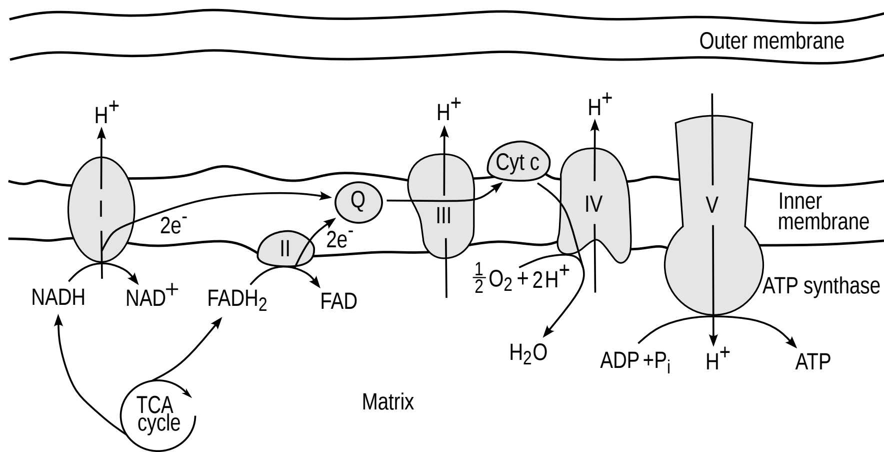

Figure 2.2 Schematic diagram of the ETC and ATP synthases that comprise the OXPHOS system. The TCA cycle supplies NADH and FADH2, each of which donate a pair of electrons to the ETC via OXPHOS complexes I and II, respectively. The electrons are then donated to ubiquinone (Q) and transferred to Cytochrome c (Cyt c) via complex III. Cytochrome c transports electrons to complex IV, where molecular oxygen is reduced to water. The transfer of electrons results in the pumping of protons (H^+) into the intermembrane space, creating an electrochemical potential. The membrane potential is dissipated by the re-entry of protons back to the matrix through the ATP synthase (complex V), which results in the production of ATP from adenosine diphosphate (ADP) and inorganic phosphate (P_i). (Nolfi-Donegan et al. 2020; Zhao et al. 2019.)

From these complexes I and II, the electrons are transferred to ubiquinone, which serves as an electron carrier to further move the electrons to OXPHOS complex III. Ubiquinone is reduced to ubiquinol, which is then oxidized by complex III, causing the electrons to continue their journey to cytochrome c and protons to be moved to the intermembrane space. OXPHOS complex IV then oxidizes cytochrome c and transfers the electrons to oxygen, which is reduced to water in aerobic cellular

respiration. Again, protons are pumped from the matrix, with half of them being used to form water and the other half transferred to the intermembrane space. (Nolfi-Donegan et al. 2020; Zhao et al. 2019.)

The generation of one water molecule is accompanied by the pumping of four, four and two protons from the matrix to the intermembrane space through complexes I, III and IV, respectively. The protons generate an electrochemical gradient known as the mitochondrial membrane potential. This gradient powers ATP synthase, also known as OXPHOS complex V, through the proton flow down the gradient back to the matrix. This results in the synthesis of ATP from adenosine diphosphate and inorganic phosphate. (Nolfi-Donegan et al. 2020; Zhao et al. 2019.) The electrochemical potential also drives other mitochondrial functions, such as the calcium ion uptake via mitochondrial Ca^{2+} uniporter (Rizzuto et al. 2012), and it can also be uncoupled to generate heat (Ricquier and Bouillaud 2000).

2.2 Mitochondrial DNA (mtDNA)

Mitochondria are believed to have originated from the engulfment of an α-protobacterium by a primordial eukaryotic cell around two billion years ago (Gray et al. 1999; Margulis 1970). During evolution, they have acquired the aforementioned additional functions in the cell, and most of the genetic material of the α-protobacterium has been lost or transferred to the nuclear genome, giving rise to nDNA sequences of mitochondrial origin (NuMTs). The mitochondrion is the only organelle in mammalian cells harbouring its genome outside of the nucleus. (Ricchetti et al. 2004; Zhang and Hewitt 1996.) What remains in human mitochondria is a double-stranded, circular mtDNA molecule consisting of 16,569 base pairs. The two strands are differentiated into heavy (H), guanine-rich and light (L), cytosine-rich strands. (Anderson et al. 1981; Andrews et al. 1999.)

MtDNA contains 37 genes encoding 13 protein subunits of the OXPHOS complexes I, III, IV and V. Complex II is encoded entirely by nDNA. The RNA components of the mitochondrial gene expression machinery, 22 tRNAs and two ribosomal RNAs (rRNAs), are also encoded by mtDNA (Figure 2.3). Seven of the 13 protein subunits contribute to the OXPHOS complex I (ND1–6 and ND4L), one to complex III (CYB), three to complex IV (CO1–3), and two to complex V (ATP6 and ATP8). However, nDNA encodes the majority of the protein subunits, as well

as the proteins required for the ETC assembly. The control region contains control elements for mtDNA transcription and replication and is considered the most variable region of mtDNA, with three polymorphism-concentrated hypervariable (HV) regions. (Chinnery and Hudson 2013.) The proteome of human mitochondria is estimated at ∼1,500 proteins, indicating that the majority of mitochondrial proteins are encoded by nDNA (Pfanner et al. 2019).

Figure 2.3 The human mitochondrial genome. It includes 37 genes, 28 on the heavy (H) strand and 9 on the light (L) strand. The genes encode 13 polypeptide components of the four OXPHOS complexes, 22 tRNAs (two-letter code) and two rRNAs (RNR1 and RNR2). The 'MT-' prefix is omitted from each gene name for clarity. The non-coding control region (CR) is considered the most variable region of mtDNA, with three polymorphism-concentrated hypervariable (HV) regions. Derived from 'Map of the human mitochondrial genome' by Emmanuel Douzery, (CC BY-SA 4.0), via Wikimedia Commons.

MtDNA has unique characteristics distinguishing it from nDNA; the genes lack introns and have no or only a few non-coding bases between them. The control region is the only significant non-coding area. (Andrews et al. 1999.) MtDNA is present in multiple copies per cell: the number varies between 100 and 10,000 copies, depending upon cell and tissue type and cellular energy demand (Chinnery and Hudson 2013). The relative quantity of mtDNA compared to nDNA, the mitochondrial copy number, declines with age and is associated with general health among the elderly (Mengel-From et al. 2014). A reduced mtDNA copy number has also been associated with an increased risk of cardiovascular disease (Ashar et al. 2017), cardiometabolic disease traits (Liu et al. 2021) and neurodegenerative disease (Yang et al. 2021).

MtDNA is replicated continuously and randomly, independent of the cell cycle (relaxed replication) (Birky 1994). A major difference is that while nDNA follows the laws of Mendelian biparental inheritance, mtDNA is inherited from the mother (Giles et al. 1980; Hutchison et al. 1974). Paternal transmission of mtDNA has also been reported (Luo et al. 2018; Schwartz and Vissing 2002), but this theory has been challenged and considered to be a technical artefact resulting from NuMTs (Lutz-Bonengel and Parson 2019; Pagnamenta et al. 2021; Rius et al. 2019).

MtDNA was the first significant part of the human genome to be sequenced, and, later, this nucleotide sequence of one individual of European descent was designated as the Cambridge Reference Sequence (Anderson et al. 1981). A corrected version was published two decades later (Andrews et al. 1999), and since then, the revised Cambridge Reference Sequence (rCRS) has been used as the standard reference sequence to annotate mtDNA in molecular anthropology, forensic science and medical genetics (Bandelt et al. 2014). The use of an European reference sequence has resulted in practical problems in some applications, and, therefore, a new reference sequence, the Reconstructed Sapiens Reference Sequence (RSRS), has been proposed (Behar et al. 2012). The RSRS maintains the same position numbering as rCRS but represents the ancestral genome of 'Mitochondrial Eve'. A third reference sequence that is used by, for example, some genotyping microarrays is an African Yoruba sequence with a length of 16,571 base pairs.

2.2.1 Variation of mtDNA

The mutation rate of mtDNA is much higher than that of nDNA, approximately 10–100 times faster than nDNA sequences with comparable function, depending on the mtDNA region (Brown et al. 1979; Neckelmann et al. 1987; Pesole et al. 1999). This is believed to result from the proximity of mtDNA to the reactive oxygen species (ROS) generated during ETC and from the less effective repair machinery than that of nDNA (Chinnery and Hudson 2013; Stewart and Chinnery 2015). Another hypothesis is that the investment of additional energy to mutation repair would impair an individual's ability to transmit genetic information to the next generation (Wallace 2007).

Mitochondrial single-nucleotide polymorphisms (mtSNPs) are most often transitions (A>G, G>A, T>C, or C>T), whereas transversions (A>C, G>T, C>G etc.) are about 15 times less frequent (Tamura and Nei 1993). When a point mutation occurs in one of the many mtDNA molecules of the cell, a mixture of wild-type and mutated mtDNA molecules is created, a state known as heteroplasmy. That is, in contrast to nDNA genetic alterations that are dichotomous, mtDNA mutations have both quantitative and quantized characteristics. When a cell with heteroplasmic mtDNA then divides, the two types of mtDNA can be distributed unequally between the daughter cells, a phenomenon known as vegetative segregation. Mutations that have occurred within approximately three human generations are usually heteroplasmic, and within those generations, the genetic drift drives the heteroplasmic mtDNA towards a pure mutant or back to a wild-type mtDNA population, a state known as homoplasmy. (Stewart and Chinnery 2015; Wallace 2007.) This rapid segregation of new mtDNA mutations to homoplasmy has been explained by relaxed replication and vegetative segregation, together with the 'mitochondrial bottleneck' hypothesis whereby the offspring receives only a small subsample of the maternal mtDNA (Cree et al. 2008; Hauswirth and Laipis 1982; Stewart and Chinnery 2021).

The high mutation rate combined with maternal inheritance could lead to an accumulation of pathogenic mtDNA mutations, ultimately resulting in mutational meltdown, a hypothesis known as Muller's ratchet (Muller 1964). However, the 'mitochondrial bottleneck', together with strong purifying selection, eliminates most pathogenic mutations from the female germ line. Purifying selection is considered a mode of natural selection by which alleles with reduced fitness or viability are lost

in a population. (Buford et al. 2018; Stewart et al. 2008.)

While point mutations may be passed down the maternal lineage, deletions are, for reasons not completely known, rarely inherited, and insertions are not thought to be pathogenic (Chinnery et al. 2004; Wei and Chinnery 2020). It is thought that most mtDNA deletions are induced by replication errors (Krishnan et al. 2008). The clinical severity of a deletion-induced disease correlates with the level and tissue distribution of the deletion, and mitochondrial dysfunction results from the loss of key OXPHOS genes (Chinnery and Hudson 2013).

Approximately one-fourth to one-third of variants within gene-encoding regions appear to be functionally important (Wallace 2010). For a pathogenic variant, the heteroplasmy level must exceed a critical biochemical threshold before a defect in mitochondrial function can be detected using established laboratory techniques. The biochemical threshold varies from mutation to mutation and from tissue to tissue, but is typically 60%–80%. However, it has been suggested that mtDNA mutations can have biological effects already at much lower heteroplasmy levels. (Stewart and Chinnery 2015; Stewart and Chinnery 2021.) As next-generation sequencing methods have been improved, low levels of heteroplasmy (0.2%–2%) have been identified in virtually all healthy individuals and the heteroplasmy level increases during life, potentially contributing to common late-onset diseases (Payne et al. 2013). A study of ∼13,000 whole-genome sequences revealed that almost half of them showed mtDNA heteroplasmy levels of greater than 1% (Wei et al. 2019). Higher heteroplasmy levels, above 10%, have been considered moderately rare in healthy individuals (Stewart and Chinnery 2021), even though a recent Genome Aggregation Database study observed that 1/250 individuals carry a pathogenic mtDNA variant with heteroplasmy of above 10% (Laricchia et al. 2022). The highest relative number of heteroplasmic variants have been detected in tissues with high metabolic activity, such as muscle and liver, and in lower numbers in blood and bone, for example. (Naue et al. 2015). From an evolutionary perspective, low-level heteroplasmy could provide a pool of potential beneficial variants to secure the best mitochondrial respiratory machinery for the offspring (Suomalainen 2019).

Pathogenic mtDNA variants fall within a continuous range, from those which are causal to monogenic disease to common mtDNA polymorphisms affecting the risk of developing noncommunicable complex human disorders (Figure 2.4). Rare and inherited pathogenic variants are usually heteroplasmic and located in protein-

coding and tRNA genes. Diseases caused by these mutations typically affect multiple organ systems, such as mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes syndrome. (Chinnery et al. 2012; Stewart and Chinnery 2015.) Some homoplasmic variants cause disease only together with certain environmental or genetic factors. An example of a such disease is Leber's hereditary optic neuropathy (LHON), resulting in optic nerve degeneration. Mainly males carrying the predisposing mutations are at an increased risk of developing the disease. Smoking and heavy alcohol intake are major risk factors for LHON, even though both LHON patients and asymptomatic mutation carriers use these substances in higher amounts, when compared to the general population. (Kirkman et al. 2009; Rabenstein et al. 2021.)

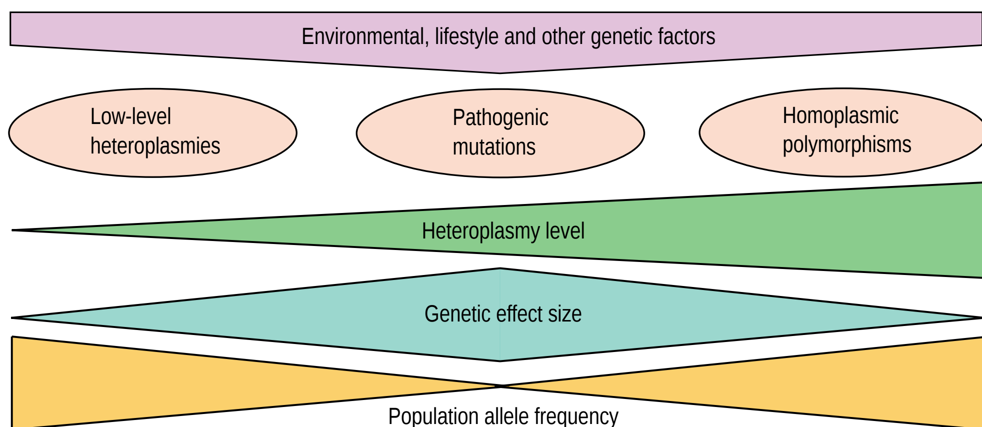

Figure 2.4 MtDNA variation and its frequency in human disease. Low-level heteroplasmy is extremely common, and the level increases during life. Intermediate-level heteroplasmy is rare in healthy individuals and may cause severe disorders affecting multiple organ systems. Common homoplasmic polymorphisms are usually neutral but may contribute to the risk of developing noncommunicable diseases. Figure adapted and modified from Wei and Chinnery 2020.

It is estimated that ~1 in 5,000 individuals have manifested disease resulting from pathogenic mtDNA mutations (Gorman et al. 2015). Approximately half of the confirmed pathogenic mutations in mtDNA are located within tRNA- and rRNA-encoding regions, despite the fact that these regions cover only about 5% of mtDNA (Lott et al. 2013). Pathogenic mutations in tRNA and rRNA genes are thought to decrease mitochondrial protein synthesis and destabilize tRNA's secondary or tertiary structures, whereas mutations in the protein-encoding genes affect specific OXPHOS complexes (Greaves et al. 2012).

2.2.2 Common variants and mitochondrial haplogroups

As humans have migrated and populated the globe, maternal inheritance and the high mutation rate of mtDNA have led to a broad range of stable population-specific, geographically isolated polymorphisms, creating phylogenetically related haplotypes. In general, approximately 5% of the mtDNA variants are present with an population allele frequency greater than 5%. Each new stable mutation has created a new branch to a complex maternal family tree reaching back to one woman, the 'mitochondrial Eve', who is postulated to have lived ∼200,000 years ago in Africa. Haplotype comparisons of humans of diverge origins have helped in shedding light on human prehistory and population movements. (Cann et al. 1987; Ingman et al. 2000; Stewart and Chinnery 2021; Wallace et al. 1999.)

The most common haplotypes define related groups of mtDNA, known as mitochondrial haplogroups. Population-based mtDNA sequencing studies have been defining new haplogroups to this day, and the current global mtDNA phylogenetic tree includes nearly 5,500 sub-haplogroups (van Oven and Kayser 2009). The geographic origin of major mitochondrial haplogroups is shown in Figure 2.5. The vast majority of European mtDNA falls into ten major haplogroups: H, I, J, K, M, T, U, V, W, and X (Torroni et al. 1996).

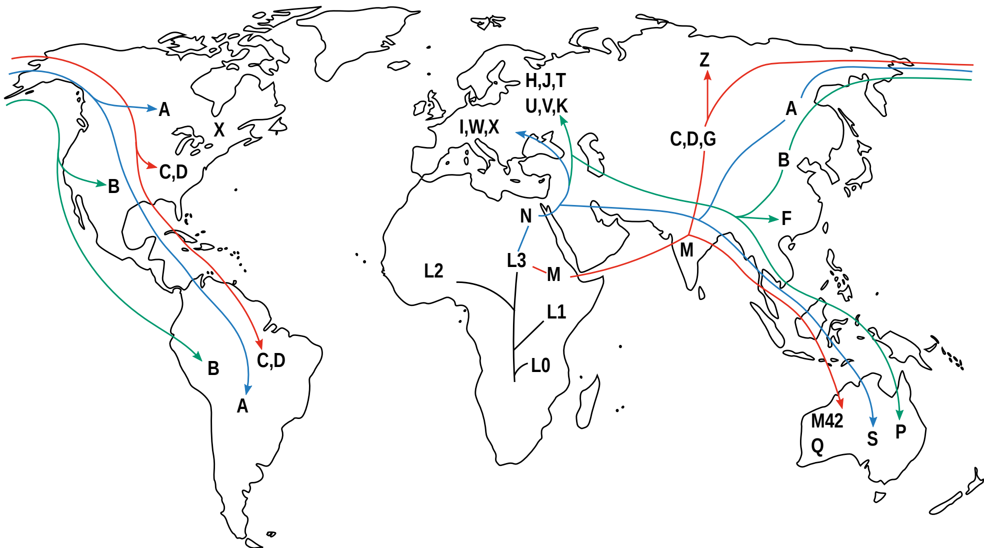

Figure 2.5 Geographic origin and distribution of mitochondrial haplogroups, based on Lott et al. 2013.

European haplogroups are also prevalent in the Finnish population, in which the most common haplogroups are H and U. These two haplogroups cover approximately 65% of the common mtDNA variation in Finland. Haplogroup H has been shown to have the highest frequency in all Finnish geographic subpopulations, except among the Sámi and those with Finnish Karelian geographic ancestry. In these populations, haplogroup U is the most frequent. In Southwest Finland, the frequency of haplogroups I, W and X is five times higher than in other subpopulations. (Hedman et al. 2007; Ingman and Gyllensten 2007.)

The haplogroup-defining variants are neutral or near-neutral from the standpoint of natural selection and have come prevalent during genetic drift (Wallace et al. 1999). However, mitochondrial haplogroups have been suggested to alter the risk of several noncommunicable diseases (Friedrich et al. 2021), including Parkinson's disease (Hudson et al. 2013), Alzheimer's disease (Santoro et al. 2010), and ischaemic stroke (Chinnery et al. 2010), for example. Among patients admitted to intensive care, haplogroup H has been associated with an increased chance of survival after severe sepsis (Baudouin et al. 2005). Taking into account that the haplogroup did not affect the risk of admission to intensive care, it could be reasoned that different haplogroups are associated with different levels of 'fitness' under external stressors (Stewart and Chinnery 2021).

It has also been suggested that the excess nonsynonymous variants have helped the Arctic populations to adapt to colder climates by allowing mitochondria to generate more heat at the expense of ATP production (Ruiz-Pesini et al. 2004), but this theory has not been generally accepted (Kivisild et al. 2005; Saxena et al. 2006). Instead, the surplus of nonsynonymous variants has been explained by the fact that, from an evolutionary viewpoint, the Arctic populations are younger and the variants have not yet been removed by the purifying selection (Stewart and Chinnery 2021).

2.3 Gene expression and epigenetics

2.3.1 DNA transcription

Gene expression is a fundamental process by which the genetic information in the DNA is used for the synthesis of functional products like proteins and RNAs. The first step of gene expression is transcription, during which a specific DNA segment is

copied into an RNA. The term transcriptome refers to the set of all RNA molecules from protein-coding messenger RNA (mRNA) to non-coding RNA. The second major step is translation, when the genetic code in mRNA is decoded in a ribosome to produce a specific polypeptide. Later, the polypeptide folds into a biologically functional protein. (Liu et al. 2016b.)

An altered gene expression profile can be considered to reflect changes in protein levels and biological processes in the human body. Analysis of transcript data helps to gain a deeper understanding of normal cell function and how changes in gene expression reflect or contribute to disease. However, many studies have concluded that there is poor correlation between mRNA and protein concentrations, indicating that transcript levels by themselves may not be sufficient to predict protein levels in many scenarios and to thus explain genotype-phenotype relationships (Liu et al. 2016b; Maier et al. 2009).

2.3.2 DNA methylation

Epigenetics refers to mechanisms affecting gene function which cannot be explained by changes in DNA sequence and which can be inherited by cell division. Epigenetic changes are mediated via modification of the chromatin structure, which varies between the condensed and transcriptionally repressed, and decondensed and transcriptionally active state. The main epigenetic mechanisms of DNA methylation (DNAm), include histone post-translational modification, small interfering RNAs and histone variants. (Margueron and Reinberg 2010.)

DNAm involves the conversion of the cytosine base in DNA into 5-methylcytosine (5mC) by the addition of a methyl group (-CH_3). In mammals, methylation mainly occurs in the regions where a cytosine nucleotide is followed by a guanine nucleotide in a linear sequence in the 5' to 3' direction, while separated by a phosphate group. These CpG sites are often clustered in high-frequency genomic regions known as CpG islands. The 'p' stands for the phosphodiester bond that joins the two nucleotides. (Matilainen et al. 2017.)

DNAm is mainly catalysed by three distinct DNA methyltransferases: DNMT1, DNMT3A, and DNMT3B. The first maintains the methylation patterns upon DNA replication, while DNMT3A and DNMT3B are mainly responsible for *de novo* methylation. The source of methyl groups used by DNMTs is S-adenosyl methionine

(SAM). It is generated through the coupling of the folate and methionine cycles in the cytosol, which in turn are sustained by the one-carbon cycle in mitochondria. The methionine cycle is also dependent on cellular ATP, indicating that mitochondrial function can regulate DNAm. DNA demethylation is carried out by three TET methylcytosine dioxygenases: TET1–3. These enzymes are catalysed and inhibited by metabolites produced in the TCA cycle. (Greenberg and Bourc'his 2019; Matilainen et al. 2017.)

The basic role of DNAm is to restrict DNA accessibility by wrapping the DNA within the chromatin structure. When the regulatory sequences of a gene, such as promoters or enhancers, are methylated, gene expression is repressed. Regardless of tissue and species, a quasi-linear inverse relationship between DNAm of the first intron and gene expression has been shown (Anastasiadi et al. 2018). It has also been demonstrated that active demethylation of several gene regulatory sequences is essential to proper gene expression (Orlanski et al. 2016). Hence, DNAm does not turn genes off but can prevent gene activation. Methylation states can also be regarded as a long-term memory of previous gene expression decisions that were mediated by transcriptional factors which might no longer be present in the cell (Dor and Cedar 2018).

Whether mtDNA itself is methylated has been under debate, and the possible functional roles remain unknown. DNMTs and TETs have been spotted in the mitochondria, but their role inside the organelle is unclear, and the presence of DNMTs is tissue-specific. Measurements of 5mC levels have suggested that this epigenetic modification occurs inside the mitochondria but at levels too low to be considered significant. Methodological studies have concluded that these methylation signals may only represent technical artefacts. Other studies claim that mtDNA is methylated differently to nDNA. (Lopes 2020; Stewart and Chinnery 2021.)

2.4 Mitochondrial control of gene expression and DNA methylation

As the majority of the mitochondria-destined proteins are encoded by nDNA, proper communication between the mitochondrial and nuclear genome is vital in order to maintain normal cellular function under different physiological and environmental

conditions. Classically, two distinct signalling pathways that link the mitochondria and nucleus have been identified: 1) anterograde signals from the nucleus to mitochondria through the activation of nuclear transcription and cytoplasmic mRNA translation, and 2) retrograde signals from the mitochondria to the nucleus for responding to metabolic conditions prevailing in the mitochondria and, finally, regulating mitochondrial functionality and metabolism. Retrograde signals are also considered a cellular adaptation to dysfunctional mitochondria, resulting from mutations in either mtDNA or nDNA-encoded mitochondrial genes. (Guha and Avadhani 2013; Matilainen et al. 2017; Quirós et al. 2016.)

Classically, retrograde signals have been considered a means for adapting to mitochondrial stress. Mitochondrial impairment has been shown to induce changes in the nuclear methylome, some of which could be reversed by the reintroduction of wild-type mitochondria to the cells (Smiraglia et al. 2008). However, recent studies have suggested that mitochondrial–nuclear communication is also a continuous process that is not governed exclusively by acute stress-inducing events (Fetterman and Ballinger 2019; Kopinski et al. 2019).

2.4.1 Mechanisms of mitochondrial–nuclear communication

Mitochondria generate a broad range of cell-specific retrograde signals through which they alter the expression of nDNA-encoded genes involved in metabolic reprogramming and stress response (English et al. 2020). An overview of the central signalling methods is shown in Figure 2.6.

Energetic response

Since mitochondria are the main site for ATP generation, ATP levels are sensitive signals relaying metabolic cues to the nucleus. Reduced ATP synthesis stimulates the adenosine monophosphate-activated protein kinase (AMPK) pathway that stimulates peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PGC1α), leading to a transcriptional adaptation to energy deficiency, and to anterograde signals that increase mitochondrial energy metabolism and biogenesis. ATP is also an allosteric inhibitor of pyruvate and isocitrate dehydrogenases acting in the TCA cycle, from which the metabolites mediate signals to the nucleus. (Garcia-Roves et al. 2008; Martínez-Reyes and Chandel 2020; Mottis et al. 2019.)

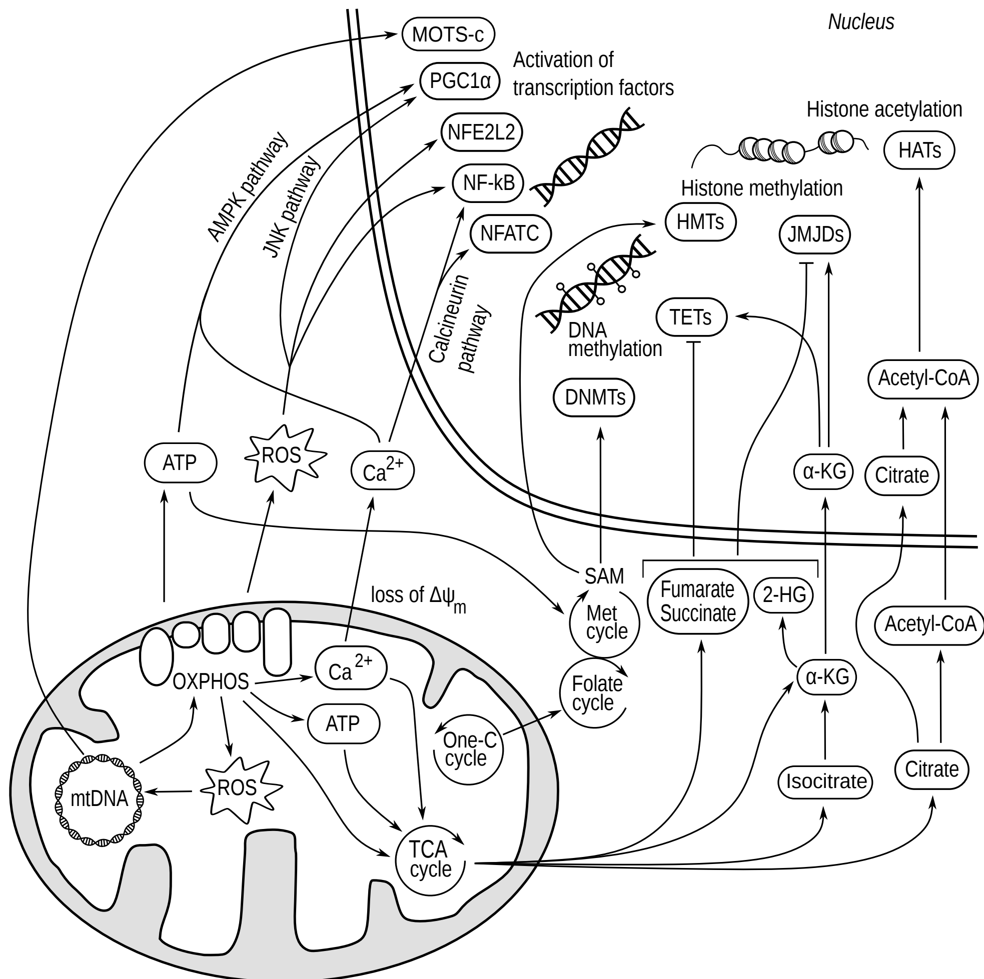

Figure 2.6 A simplified overview of mitochondrial–nuclear communication. Increased activation of nuclear transcription factors may result from a decrease in ATP levels, elevated ROS levels, or from the release of Ca^{2+} from the mitochondrial matrix. Mutations in mtDNA may contribute to these signalling pathways via defects in OXPHOS that affect ATP and ROS levels and cause the loss of the mitochondrial membrane potential ($\Delta\Psi_m$). Mitochondrial-derived peptide MOTS-c translocates to the nucleus to regulate nuclear gene expression in response to metabolic stress. Metabolites produced in the TCA cycle affect gene expression via chromatin modifications and DNAm. Nuclear acetyl-CoA fuels histone acetyltransferases (HATs). α-Ketoglutarate (α-KG) serves as a cofactor for histone and DNA demethylases (JMJDs and TETs), whereas succinate, fumarate and 2-hydroxyglutarate (2-HG) inhibit the histone and DNA demethylases. Mitochondrial Ca^{2+} regulates TCA dehydrogenase activity, which provides a link between mtDNA mutations, TCA cycle metabolites and nuclear gene expression. Mitochondrial one-carbon (One-C) cycle, coupled with the cytosolic folate and methionine (Met) cycles generates SAM which is the source of methyl groups used by histone and DNA methytransferases (HMTs and DNMTs). The synthesis of SAM also depends on ATP, also indicating that mitochondrial function can regulate histone and DNA methylation. (English et al. 2020; Martínez-Reyes and Chandel 2020; Matilainen et al. 2017; Mottis et al. 2019; Quirós et al. 2016.)

Ca^{2+}-dependent signalling

Driven by the electrochemical gradient across the inner membrane, Ca^{2+} accumulates into the mitochondrial matrix, and mitochondria play a fundamental role in regulating the levels of intracellular Ca^{2+} (Rizzuto et al. 2012). Mitochondrial stress, such as mutations in mtDNA or the disruption of OXPHOS complexes, triggers the loss of the membrane potential, which leads to the release of Ca^{2+} into the cytoplasm. Elevated levels of free intracellular Ca^{2+}, in turn, activate the calcineurin phosphatase, which activates nuclear factor kB (NF-kB) and nuclear factor of activated T cells (NFATC). The activation of these transcription factors results not only in cellular adaptation to a higher Ca^{2+} level but also in responses affecting insulin signalling, glucose metabolism and cell proliferation. The release of Ca^{2+} may also activate the AMPK pathway. (Guha and Avadhani 2013; Quirós et al. 2016.) Mitochondrial matrix Ca^{2+} also regulates the activity of TCA cycle dehydrogenases (English et al. 2020). The causal relationship between calcium signalling and mitochondrial dysfunction has been validated in a study in which the chelation of cytosolic calcium abolished the downstream signalling, leading to a decrease in ATP levels (Luo et al. 1997).

Metabolites

Recent findings have suggested that metabolic reprogramming is the main mechanism linking mtDNA variation with changes in the nuclear transcriptome and epigenome (Fetterman and Ballinger 2019; Kopinski et al. 2019). Much of the metabolites that mediate epigenetic modifications are intermediates from the TCA cycle. They serve as cofactors and substrates to enzymes that contribute to epigenetic modifications of DNA or histones. The relative amounts of different metabolites reflect nutrient availability, metabolic demand, cellular redox state and mitochondrial function. Specifically, acetyl-coenzyme A (acetyl-CoA) and α-ketoglutarate (α-KG) have been recognized as key metabolites for mediating epigenetic modifications. The TCA cycle and ETC are tightly coupled, since the complexes I and II replenish NAD^+ and FAD, respectively, that are required for the TCA cycle to keep functioning (Hertzog Santos 2021; Martínez-Reyes and Chandel 2020.)

Acetyl-CoA, which is produced from various sources in multiple compartments, enters the TCA cycle to produce citrate. Citrate can continue in the TCA cycle, or

it can travel outside the mitochondria, both into the cytosol and the nucleus, where it is converted back into acetyl-CoA. The latter pathway takes place especially under conditions of carbohydrate or glucose excess. Nuclear acetyl-CoA is an essential regulator of lipid and sterol synthesis and histone acetylation, which modulate nuclear gene expression profiles. (English et al. 2020; Martínez-Reyes and Chandel 2020.)

α-KG is generated in the TCA cycle from isocitrate by isocitrate dehydrogenase 1. α-KG diffuses into the nucleus, where it serves as cofactor for histone and DNA demethylases (Jumonji C domain-containing protein [JMJD] families and TETs, respectively). An α-KG-derived metabolite, 2-hydroxyglutarate (2-HG), along with other TCA cycle intermediates, succinate and fumarate, can also modify the epigenetic landscape by inhibiting the TET and JMJD enzymes. (Campbell and Wellen 2018; Martínez-Reyes and Chandel 2020; Mottis et al. 2019.) For example, a high intracellular α-KG to succinate ratio maintains the pluripotency of embryonic cells (Carey et al. 2015).

Reactive oxygen species

Under normal conditions, 0.2%–2% of the electrons in the ETC leak out from the process and react with oxygen to produce ROS. While ROS produced during ETC are usually associated with their damage-promoting effects at their higher concentrations, they have distinct signalling roles under physiological levels. Modest increases in ROS levels activate transcription factors such as nuclear factor erythroid 2-related factor 2 (NFE2L2), which increase the expression of antioxidant response elements. ROS may also induce the expression of genes involved in OXPHOS by activating the c-Jun N-terminal kinase (JNK) pathway or promote the uncoupling of the mitochondrial membrane potential via the AMPK pathway. (Quirós et al. 2016; Shadel and Horvath 2015.) Mitochondrial ROS also mediate other key signals to control, for example, immunity, hypoxic responses, beneficial exercise-related outcomes and other signals that promote health and longevity (Mottis et al. 2019; Ristow and Schmeisser 2014; Ristow et al. 2009). In cancer cells, ROS have been shown to promote the transcriptional activation of the NF-kB (Formentini et al. 2012).

The aforementioned molecules and their co-mediators are all secondary metabolites, transient molecules or nucleous-encoded proteins. Over the past decades, new regulatory proteins encoded as short open reading frames in the mtDNA have been identified and there is emerging evidence that these mitochondria-derived peptides (MDPs) have diverse biological roles in cell metabolism (Kim et al. 2017; Miller et al. 2020). Humanin was the first identified MDP and is encoded by the 16S ribosomal RNA gene (*MT-RNR2*). It exhibits protective effects in several cell types in response to cellular stress through regulating various signalling mechanisms, such as the JNK pathway. (Hazafa et al. 2021.)

MOTS-c (mitochondrial open reading frame of the 12S rRNA-c) is the first mitochondria-encoded MDP targeted at the nucleus to directly regulate nuclear gene expression (Kim et al. 2018). MOTS-c expression is age-dependent, it is induced by physical exercise in circulation and skeletal muscle, and MOTS-c treatment can even increase physical capacity and health in mice (Reynolds et al. 2021). In individuals with T2DM, the circulating MOTS-c levels are decreased as compared to healthy controls (Ramanjaneya et al. 2019). Thus, MOTS-c could be considered a peptide hormone by which mitochondria generate a signal to shift energy demands and expenditure (Zarse and Ristow 2015).

2.4.2 Role of mtDNA variants

The mechanisms through which mtDNA haplogroups and variants modify disease susceptibility are not completely understood, partly due to the heterogeneous nuclear genomic backgrounds of different individuals. Transmitochondrial cytoplasmic hybrid (cybrid) models have allowed more detailed investigations of the molecular and biological functional consequences resulting from mtDNA variation. These *in vitro* cybrid cell lines are created by fusing mitochondria-free cells with mitochondria-rich platelets from different individuals so that the resultant cells have identical nuclei but varying mtDNA. For example, compared to haplogroup H cybrids, haplogroup L cybrids have shown lower expression levels of complement pathway and innate immunity genes and increased levels of inflammation-related signalling genes, all of which are critical in human diseases (Kenney et al. 2014b).

The most common pathogenic heteroplasmic mtDNA variant is the m.3243A>G mutation in the *MT-TL1* gene (Stewart and Chinnery 2021). Analysis of m.3243A>G cybrids with varying heteroplasmy levels revealed that changes in heteroplasmy impact mitochondrial function. More importantly, metabolomic and histone modification analyses provided evidence that changing levels of m.3243A>G heteroplasmy affect histone modification via mitochondrial metabolites, thus changing the nuclear epigenome and transcriptome. (Kopinski et al. 2019.)

Another platform for testing the contributions of mtDNA variation while accounting for nDNA heterogeneity is achieved by the use of mitochondrial–nuclear exchange (MNX) mice. In these mouse models, nDNA from one mouse strain is combined with mtDNA from a different mouse strain. (Fetterman et al. 2013.) In MNX mice, the mtDNA background altered the nuclear expression in adipose tissue in response to a high-fat diet, irrespective of the nuclear background. The altered transcriptional response was associated with genes containing binding sites for transcription factors that are known to be ROS-sensitive. This implies that mtDNA variants may modulate ROS levels through the ETC complexes in response to metabolic stress and thereby activate signalling pathways. (Dunham-Snary et al. 2018.) Another study using MNX mice has also shown that mtDNA variation affects DNAm and gene expression in brain tissue (Vivian et al. 2017).

Studies examining the associations between mtDNA variants and nuclear or mitochondrial gene expression on a population level are scarce. Haplogroup-specific mtDNA-encoded gene expression patterns have been demonstrated in a study including $\sim$450 individuals with varying ethnic backgrounds (Cohen et al. 2016). The different expression patterns are thought to result from co-expressed nDNA-encoded mitochondrial RNA-binding proteins and differences in the RNA stability of the mature transcripts (Rorbach and Minczuk 2012).

In a family-based GWAS including $\sim$850 Australian individuals of mainly Anglo-Celtic origin, the associations of 78 common mtDNA variants with $\sim$47,000 gene expression probes in peripheral blood were examined. The study identified 15 significant associations, corresponding to five unique genes on the nuclear and mitochondrial genomes, three of which were nDNA-encoded. Some of the significant results were also replicated in an independent European cohort. The biological relevance of the associations remained unclear, and they do not necessarily imply a causal relation. However, the associations may equally represent the nuclear transcriptional

response to altered cellular homeostasis. (Kassam et al. 2016.) On a population level, there are no published association studies examining the role of mitochondrial haplogroups in nuclear gene expression.

As previously discussed, many of the mtDNA-induced transcriptome-affecting changes take place via DNAm, and there are several studies suggesting that mtDNA variants have a distinct role in DNAm. A study on murine embryonic stem cells has demonstrated that mtDNA haplotypes influence differential patterns in nuclear gene expression (Lee et al. 2017). Haplogroup J has been consistently associated with higher methylation levels in human retinal cell cybrids and articular cartilage when compared to haplogroup H (Atilano et al. 2015; Cortés-Pereira et al. 2019). Analysis of human peripheral blood has shown that the global DNAm levels in haplogroup J carriers are also higher when compared to subjects carrying haplogroups H, U, X and T (Bellizzi et al. 2012).

Cells harbouring haplogroup J have shown lower levels of ATP compared to other major haplogroups (Bellizzi et al. 2012; Fernández-Moreno et al. 2017). This could be explained by the fact that many variants clustered in haplogroup J fall within OXPHOS complexes I and III, which results in a decrease in oxygen consumption and a partial uncoupling of OXPHOS (Arning et al. 2010; Marcuello et al. 2009). In turn, ATP levels have been thought to affect *MAT1A* gene transcription. The methionine adenosyltransferase encoded by this gene catalyses the generation of SAM which is the source of the methyl groups used by DNMTs. (Bellizzi et al. 2012.)

Low levels of ROS have correlated with high DNAm in haplogroup J cybrids, whereas high levels of ROS have correlated with high methylation levels in mtDNA-depleted cells (Bellizzi et al. 2012). This suggests that ROS is not a predictor of DNAm, although it should be noted that the depletion of mtDNA is a highly artificial state.

In addition, GWASs have identified numerous DNAm quantitative trait loci (QTL) in the nuclear genome (Huan et al. 2019; Lemire et al. 2015; Min et al. 2021), and variation in the mtDNA copy number has also been associated with DNAm levels (Wang et al. 2022). However, these studies have not investigated the role of mtDNA variants and cohort-level mitochondrial GWASs of DNAm are lacking.

2.4.3 Sexual dimorphism

Since mtDNA is maternally inherited, it is not, in theory, subjected to natural selection in males. This hypothesis of a sex-specific selective sieve, colloquially known as the 'mother's curse', suggests that variants that are male-harming but neutral or beneficial for females may accumulate in the mitochondrial genome (Dowling and Adrian 2019; Frank and Hurst 1996; Gemmell et al. 2004). Classic examples of such variants are those affecting male fertility via sperm motility and morphology (Holyoake et al. 2001; Ruiz-Pesini et al. 2000), although the evidence for maternally driven selection acting in male fertility is equivocal (Mossman et al. 2012; Pereira et al. 2007). Another example suggested to support the mother's curse hypothesis is LHON, a mitochondrial disease that has a male-biased prevalence: men represent approximately 80% of cases, and 10% of women and 50% of men carrying the predisposing variant develop visual failure (Man et al. 2003; Milot et al. 2017). However, a mitochondrial genetic background cannot fully explain the sex bias in the disease (Chinnery and Schon 2003).

There is some evidence in *Drosophila melanogaster* that mtDNA variants have sex-specific effects on nuclear gene expression; the mitochondrial genetic effects were strong in males but weak in magnitude and negligible in number in females (Innocenti et al. 2011). Notably, the male-biased transcriptional hotspots were located in the testes and accessory glands. Another *Drosophila* study did not find any significant mtDNA effects targeting male-limited tissues (Mossman et al. 2016). In the same study, however, mtDNA variants exerted major effects in the nuclear transcriptome in females, whereas in males, the effects of mtDNA variation were considerably smaller. In fruit flies, mtDNA haplotypes also have male-biased effects on the metabolic rate measured by the production of carbon dioxide (Nagarajan-Radha et al. 2020).

The hypothesis of sexual dimorphic gene expression resulting from mtDNA variation has also been tested in humans. In the same mitochondrial GWAS mentioned earlier, association analyses were performed separately for males and females, and the differences in the effect sizes for mitochondrial genotypes were examined. There was no evidence of sex-specific effects. (Kassam et al. 2016.) It is possible, however, that the sexually dimorphic effects are mediated by other mtDNA variants not included in that study.

Sex-specific differences in DNAm patterns have been identified in peripheral blood and replicated in independent cohorts. The sex-associated methylation also has functional effects, a small number of differentially expressed genes were associated with CpGs sites showing sex-specific heterogeneity. (Singmann et al. 2015; van Dongen et al. 2016.) It is also possible that sex differences may mediate sex discordance in phenotypic traits (Khramtsova et al. 2019). The maternal inheritance of mtDNA could also contribute to the sexual asymmetry in DNAm, but there are no studies testing this hypothesis.

2.5 Cardiometabolic risk factors and mtDNA variants

The aetiology of cardiometabolic disorders, such as cardiovascular disease and T2DM, is represented by a cluster of interrelated risk factors, mainly elevated blood pressure (BP), fasting plasma glucose (FPG), dyslipidaemia and abdominal obesity. Cardiometabolic disorders are generally polygenic and multifactorial traits, involving the interaction of genetic, environmental and lifestyle factors. Because mitochondria lie at an interface between the bioenergetic processes of the human body, they are potent mediators of common metabolic disorders.

2.5.1 Diabetes-related changes in gene expression and DNA methylation

Prediabetes is defined by glycaemic variables that are higher than normal, but lower than T2DM thresholds. It is a major risk factor for developing T2DM; each year, 5%–10% of individuals with prediabetes will progress to T2DM, with the same proportion converting back to normoglycaemia. The prevalence of prediabetes is increasing worldwide, and it is estimated that 587 million individuals will have prediabetes by 2045. Individuals with prediabetes may already have end-organ damage that is traditionally thought to be a complication of T2DM, such as nephropathy, neuropathy, retinopathy and macrovascular disease. (Hostalek 2019; Tabák et al. 2012.)

The aetiology of T2DM has a strong genetic component, but studies have provided conflicting answers regarding the role of mtDNA variation in prediabetes and T2DM. In European populations, mtDNA variants are unlikely to play a major role

in the risk of developing the disorders (Achilli et al. 2011; Chinnery et al. 2005; Chinnery et al. 2007; Saxena et al. 2006). In Korean and Japanese populations, haplogroup N9a was associated with a decreased risk of T2DM, while haplogroups D5 and F were associated with an increased risk of T2DM (Fuku et al. 2007). However, the reported protective effect of N9a has been challenged (Takasaki 2009), and the same haplogroup has also been linked with an increased risk of T2DM (Fang et al. 2018). A recent multi-centre study identified one common variant in the hypervariable region associated with FPG levels (Kraja et al. 2019).

Even though mtDNA variation does not seem to affect the susceptibility to T2DM in the European population, it may still have functional consequences strong enough to occur on a molecular level. This hypothesis is backed up by a study which examined mitochondrial cybrids harbouring haplogroups N9a, D5 and F and which found that the gene expression patterns of the cybrid cells correlated with the susceptibility to develop T2DM (Hwang et al. 2011).

Epigenome-wide DNAm changes have been associated with T2DM (Juvinao-Quintero et al. 2021; Kriebel et al. 2016; Walaszczyk et al. 2018), and there is also evidence that epigenetic changes are likely to be an early process that may already occur during the prediabetic stage (Matsha et al. 2016). The cross-talk between mtDNA and DNAm in the setting of T2DM or prediabetes, however, is not known. It should also be noted that the epigenome can both affect the disease and be affected by it, often making the direction of causation obscure (Michels et al. 2013).

2.5.2 Hypertension-associated variants

In the industrialised countries, the risk of becoming hypertensive (BP $>$ 140/90 mmHg) during a lifetime exceeds 90% (Messerli et al. 2007). Worldwide, hypertension is estimated to cause over 9 million premature deaths per year (Lim et al. 2012). Genetic variance has been estimated to account for 19%–56% of systolic blood pressure (SBP) variation, 37%–52% of diastolic blood pressure (DBP) variation and up to 82% of mean arterial pressure (MAP) variation (Hottenga et al. 2006; Steves et al. 2012). Most of the hypertension-associated genetic variants have been identified in the nuclear genome (e.g. Ehret et al. 2011; Liu et al. 2016a; Tragante et al. 2014), and few studies have specifically focused on the mitochondrial genome.

Mutations in mtDNA are thought to increase BP via increased ROS production

(Ding et al. 2013). This hypothesis is strengthened by a study in which the downregulation of the mtDNA-encoded cytochrome-b (mt-CytB) gene was observed in hypertensive rats. This gene seems to contribute directly to mitochondrial ROS production. An injection of microRNA-21 counteracted the mt-CytB downregulation and lowered the BP. (Li et al. 2016.) Another proposed mechanism is the dysregulated Ca^{2+} uptake into the endothelial mitochondria (Chen et al. 2016). Mitochondrial Ca^{2+} modulates the generation of ROS and nitric oxide, which have an effect on vascular and sympathetic tone (Brandes 2014).

A mutational hotspot for pathogenic mtDNA variants associated with hypertension is located in the tRNA coding regions (Ding et al. 2013). However, the majority of associations have been identified in case reports examining Han Chinese families (Chen et al. 2012; Teng et al. 2012; Wang et al. 2011), implying that these variants are rare on a population level.

Some evidence of the role of mtDNA variants has also been gleaned from well-established cohorts. In the Framingham Heart Study, the analysis of over 7,200 participants of European descent identified that a rare (allele frequency 0.6%) nonsynonymous variant m.5913G>A in the *MT-CO1* gene was associated with higher SBP (Liu et al. 2012). In a two-cohort meta-analysis of older North American individuals, in which the variants were sequenced, two (m.3197T>C and m.15924A>G) and five (m.93A>G, m.12705C>T, m.16172T>C, m.16183A>C, and m.16189T>C) variants associated with variation in SBP and MAP in white and black participants, respectively. In addition, a rare variant analysis testing the aggregate effects of variants within a specified region identified significant pooled effects across all tRNA regions. (Buford et al. 2018.) These associations have not been replicated in other studies.

Negative results from population-level analyses have also been published. Sixty-four tagging mtDNA variants that efficiently capture all common European variation (except the hypervariable region) did not associate with BP in a study consisting of over 2,000 individuals (Saxena et al. 2006). A smaller sequencing study with ~360 participants utilized the predicted pathogenicity of the protein coding variants but did not find a significant role for mtDNA variation in association with BP levels (Venter et al. 2017). A lexical tree analysis of over 2,800 hypertensive individuals and 5,600 controls with phylogenetically related mtDNA variants in a European population did not identify significant relationships with hypertension (Hudson et al. 2014).

3 AIMS OF THE STUDY

The aim of this study was to investigate the mitochondrial genetic determinants of peripheral blood gene expression, DNA methylation and blood pressure.

The specific aims were to:

1. examine the associations of mtDNA variants with genome-wide transcriptome and epigenome-wide DNAm profiles in peripheral blood by means of a GWAS (studies I and II).
2. study whether the aforementioned associations are affected by the onset of prediabetes (I and II).
3. study the role of mtDNA variants in BP variation by means of a GWAS (III).
4. examine whether the associations with the transcriptome, DNAm or BP show sexual dimorphism (I–III).

4 MATERIALS AND METHODS

4.1 Study cohorts

4.1.1 The Young Finns Study (I–III)

The Young Finns Study (YFS) is a Finnish longitudinal population study on the evolution of cardiovascular risk factors from childhood to adulthood (Raitakari et al. 2008). The study began in 1980, when 3,596 children and adolescents in six age cohorts (3–18 years) randomly chosen from the national register across the five university hospital catchment areas participated in the baseline studies. During the follow-ups in 2007 and 2011, 2,204 and 2,060 individuals, respectively, participated in the examinations. The blood samples for mtDNA sequencing were drawn in 2007, and other data used in this study were obtained in 2011.

4.1.2 The Ludwigshafen Risk and Cardiovascular Health study (II)

The Ludwigshafen Risk and Cardiovascular Health (LURIC) study consists of 3,316 patients of German ancestry referred for coronary angiography at a tertiary care centre in Southwestern Germany between June 1997 and May 2001 (Winkelmann et al. 2001). The clinical indications for angiography were chest pain or non-invasive tests consistent with myocardial ischaemia. Patients with any acute illness other than acute coronary syndrome, any predominant non-cardiac disease and/or a history of malignancy within the past five years were excluded from the study.

4.1.3 The Finnish Cardiovascular Study (III)

The Finnish Cardiovascular Study (FINCAVAS) participant pool consists of patients who underwent an exercise stress test using a bicycle ergometer at Tampere University Hospital between October 2001 and the end of 2008 and who were willing to participate (Nieminen et al. 2006). A total of 4,068 participants completed a technically successful exercise test. The main indications for the exercise test were a suspicion of coronary heart disease (frequency 46%), the evaluation of work capacity (26%), testing for vulnerability to arrhythmia during exercise (25%), and assessing the adequacy of coronary heart disease treatment (13%); some patients had more than one indication.

4.1.4 Ethical considerations (I–III)

All participants in the study cohorts gave their written informed consent, and the studies were conducted in accordance with the Declaration of Helsinki. The YFS was approved by the ethical committee of the Hospital District of Southwest Finland and by local ethical committees, and the study protocol of each study phase corresponded with the proposal by the World Health Organization. The LURIC study was approved by the ethics committee of the State Chamber of Physicians of Rhineland-Palatinate. The FINCAVAS protocol was approved by the Ethical Committee of the Pirkanmaa Hospital District.

4.2 Identification of mtDNA variants

4.2.1 Next-generation sequencing (I–III)

In the YFS, mtDNA variants were determined by next-generation sequencing (NGS). Genomic DNA concentrations were measured from peripheral blood samples ($n = 1,807$), and mtDNA was amplified from the genomic DNA and further processed into Illumina sequencing-compatible libraries. The sequencing process has been described in detail in original publication III. As a result, 1,658 samples were available for further analysis.

Study I

NGS data were processed and variants were called at the Institute for Molecular Medicine Finland (FIMM) using an in-house-developed bioinformatics pipeline (Sulonen et al. 2011). The following quality control (QC) filters were applied after variant calling: sample missingness > 0.10, variant missingness > 0.05, and variant allele frequency (VAF) < 0.01. Heteroplasmic genotypes were set to missing.

Studies II and III

NGS data were first aligned with the rCRS (Andrews et al. 1999) by using BWA-MEM v. 0.7.17 (Li and Durbin 2009) and SAMtools v. 1.8 (Li et al. 2009). Variants were called with Mutserve v. 1.2.1, a stand-alone version of the web tool mtDNA-server (Weissensteiner et al. 2016a), with the default thresholds for mapping, base and alignment quality scores.

The minimum heteroplasmy level was set to 5% – we defined sites with a heteroplasmy level below this threshold as homoplasmic wild-type alleles and sites with a heteroplasmy level above 95% as homoplasmic variants. Genotypes for heteroplasmic variants overlapping with any NuMTs (Dayama et al. 2014) were set to missing. The heteroplasmy rate was low at most sites: for 99% of the sites, the number of heteroplasmic samples was three or fewer. Mutserve identified variants in 1,365 different nucleotide positions from 1,657 samples. The average sequencing coverages per sample and per mtSNP were 497 and 525, respectively.

4.2.2 Genotyping arrays (II and III)

In the LURIC study, genomic DNA was extracted from peripheral blood and the mtSNPs were genotyped using the Illumina HumanExome-12 v1.2 BeadChip ($n = 1,981$) and the Illumina 200k MetaboChip ($n = 3,150$) microarrays. Samples with a call rate of < 0.95, sex mismatch and cryptic relatedness ($\hat{\pi} > 0.2$) were removed using PLINK v.1.90b6.21. Variants with an allele frequency of < 0.005 or a call rate of < 0.95 were also excluded. Heterozygous genotypes, possibly due to mitochondrial heteroplasmy, were coded as missing.

In the FINCAVAS, genomic DNA was extracted from peripheral blood leukocytes, and the Illumina Cardio-MetaboChip and HumanCoreExome-12 v1.1 arrays

were applied for genotyping the mtSNPs from 2,824 and 1,032 samples, respectively. Samples with call rate of < 0.95, excess heterozygosity, cryptic relatedness ($\widehat{\pi} > 0.2$) and sex mismatch, as well as genetic outliers based on multi-dimensional scaling plots, were removed. Variants with a call rate of < 0.95 or an allele frequency of < 0.01 were removed. Heterozygous genotypes were coded as missing.

4.2.3 Haplogroup classification (I and II)

Mitochondrial haplogroups were determined using HaploGrep 2 (Weissensteiner et al. 2016b) with Phylotree builds 16 (van Oven and Kayser 2009) and 17 (van Oven 2015) in studies I and II, respectively. In the YFS, samples with a haplogroup quality score of ≥ 0.90 were included. Since some LURIC study participants were genotyped with both Illumina microarrays, haplogroups were included based on two criteria: 1) a quality score of ≥ 0.90 in at least one genotyping batch, or 2) a quality score of ≥ 0.80 and the same major haplogroup assigned in both arrays. Haplogroups with a frequency of <1% were excluded, leaving nine major haplogroups for further analyses.

4.3 Genome-wide transcriptome analysis (I and II)

RNA isolation was performed from peripheral blood samples, and the transcript levels were analysed with an Illumina HumanHT-12 v4 Expression BeadChip containing 47,231 expression and 770 control probes. The transcripts detected (detection p-value < 0.01) in less than 5% of samples and samples with less than 6,000 significantly detected expression probes (detection p-value < 0.01) were rejected. The transcriptome data was processed in R using a non-parametric background correction, followed by quantile normalization with control and expression probes, a log2 transformation, zero-centering, and a rank-based inverse normal transformation. Based on RPS4Y1–2 and XIST mRNA levels on the Y and X chromosomes, respectively, samples with mismatch between the recorded and predicted sex were excluded. After QC, expression data were available for 19,637 transcription probes and 1,650 samples.

4.4 Epigenome-wide DNA methylation analysis (II)

Genomic DNA was extracted from peripheral blood samples by using standardized methods. DNAm levels were quantified using the Illumina Infinium MethylationEPIC BeadChip that covers over 850,000 methylation sites across the nDNA.

In the YFS, data were processed using the minfi Bioconductor package in R (Fortin et al. 2017). After QC, 769,683 autosomal and 17,334 X-chromosomal CpGs from 1,529 samples remained for further analyses. In the LURIC study, QC was performed using the CPACOR pipeline (Lehne et al. 2015), resulting in 795,619 autosomal and 18,138 X-chromosomal CpGs from 2,423 samples to be included for further examination. The QC procedures for both cohorts have been described in detail in original publication II.

Beta values (ranging between 0 [no methylation] and 1 [full methylation]) were calculated according to the equation $\beta = \mathrm{M}/(\mathrm{M}+\mathrm{U}+100)$, where M and U denote the methylated and the unmethylated signal, respectively.

4.5 Definition of clinical phenotypes

4.5.1 Prediabetes (I and II)

Venous blood samples were drawn after an overnight fast. FPG and glycated haemoglobin (HbA1c) were determined using standard laboratory techniques. The classification of prediabetes was based on the criteria of the American Diabetes Association (American Diabetes Association 2014). Individuals with prediabetes were defined as having an FPG level of 5.6–6.9 mmol/l, a two-hour plasma glucose level of 7.8–11.0 mmol/l during a 75-g oral glucose tolerance test (OGTT), or an HbA1c level of 39–47 mmol/mol without a diagnosis of T2DM. Participants diagnosed with type 1 diabetes mellitus were also excluded. OGTTs were performed only for the LURIC participants.

4.5.2 Smoking (II)

In the YFS, the smoking history was self-reported and assigned to six categories based on smoking frequency (active smoker or at least once a day, once a week or more often but not daily, less often than once a week, attempts to quit, has quit, and has never smoked).

In the LURIC study, participants were categorized into five groups: heavy smokers (defined as smoking ≥ 20 cigarettes per day), light smokers, former smokers that quit smoking less than 10 years ago, former smokers that quit smoking more than 10 years ago and never-smokers. Smoking status was additionally verified by measuring the serum cotinine concentration, and we used a cut-off value of 15 µg/l (Benowitz et al. 2020) to reclassify self-reported non- or ex-smokers as active smokers.

4.5.3 Blood pressure (III)

In the YFS, BP was determined as the average of three measurements taken at two-minute intervals in a sitting position from the right brachial artery with a random zero sphygmomanometer. The FINCAVAS participants lay in the supine position for 10 minutes, after which BP was measured once by an experienced nurse using a brachial cuff. In both cohorts, the Korotkoff method was used.

The observed BP levels were adjusted for antihypertensive medication usage. The medications were self-reported by the study participants, the duration of treatment was not known, and adherence was not assessed. Adjusted SBP was calculated by increasing the recorded measure by 8, 14 and 20 mmHg for 1, 2 and ≥ 3 medication classes taken, respectively. DBP measurements were adjusted similarly by increasing the recorded measure by 4, 10 or 16 mmHg. (Cui et al. 2003.) Adjusted MAP was calculated from the adjusted SBP and DBP values as $\text{MAP} = \text{DBP} + \frac{1}{3} \times (\text{SBP} - \text{DBP})$.

4.6 Statistical methods

4.6.1 Association analyses on the genome-wide transcriptome (I)

Transcriptome, mtDNA and clinical data were available for 955 YFS participants, and 199 mtSNPs with a VAF of ≥ 0.01 were analysed under the R environment. The expression levels of 19,637 probes were modelled as a linear function of the presence (coded as 1) or absence (coded as 0) of the variant allele. The model was adjusted with age, sex, body mass index (BMI), the first 11 principal components (PCs) derived from the transcriptome data, and the first two PCs derived from the mtDNA data. Significance was defined as $p < 1.28 \times 10^{-8}$ (0.05/199/19,637).

PC analysis is a method by which a set of variables is transformed into a smaller set of PCs while preserving as much of the variation contained in the original variables as possible. The use of mtDNA PCs as covariates has been demonstrated to be a robust method to adjust for population stratification in mitochondrial association studies. In addition, the use of mitochondrial PCs effectively removes false-positive associations but does not cause a loss of power in detecting true associations. (Biffi et al. 2010; Miller et al. 2019.) Logistic PC analysis was performed on homoplasmic genotypes passing QC and with a VAF of ≥ 0.01 using the logisticPCA package (Landgraf and Lee 2020). The number of mitochondrial PCs for the single variant analyses was selected so that the median χ^2-based genomic inflation factor (λ_G) was as close to one as possible.

The concept behind λ_G is that, with the exception of a small number of SNPs being truly associated with a trait, the majority of the association test statistics will follow the normal distribution under the null hypothesis. However, population heterogeneity, cryptic relatedness and genotyping errors may cause inflation of the test statistics and increase the number of false-positive findings. This possible inflation may be detected and controlled by calculating the λ_G. (Yang et al. 2011.) In GWASs, values of $\lambda_G < 1.05$ are generally considered in the regime of minimal inflation (Price et al. 2010).

For all mtSNPs associating with a probe, pairwise linkage disequilibrium (LD) was quantified as squared Pearson correlation r^2. In order to identify the independent signals, variants with $r^2 > 0.30$ were subjected to pairwise conditional analysis by using the same linear model as in the full analysis but additionally conditioning

for one significant variant at a time. We applied a variant-specific Bonferroni correction (i.e. correction for the number of pairwise analyses made for each variant) to account for multiple testing.

Analysis of covariance was used to flag genes for those showing differential expression between haplogroups in 934 individuals. MtDNA PCs were excluded from the covariates, since they were strongly correlated with the haplogroups. All genes with a $p < 2.55 \times 10^{-6}$ (0.05/19,637) were compared using Tukey's honest significant difference test to confirm the between-haplogroup differences. A Tukey-adjusted p-value of < 0.05 was considered significant.

Sex-specific effects of variants on gene expression were tested by applying the same linear model as described above to males and females separately. Differences in effect sizes were compared by random-effect meta-analysis by using the MetaDE package (Wang et al. 2012). Heterogeneity was examined by Cochran's Q test with the corresponding p-value. A significant p-value suggests that there is a difference in the effect sizes between the sexes. The number of variants tested was 156, because for some variants, the allele frequency was < 0.01 in either males or females. Significance was then defined as $p < 1.63 \times 10^{-8}$ (0.05/156/19,637).

The effect of prediabetes was studied similarly by applying the linear model separately to individuals with prediabetes and normoglycaemic controls. The number of variants included was now 127, resulting in a significance in random-effect meta-analysis defined as $p < 2.00 \times 10^{-8}$ (0.05/127/19,637).

All significant probes from the above-mentioned analyses were tested for cross-hybridization with sequences other than the target transcript, using algorithm blastn from BLAST (Altschul et al. 1990), with an emphasis on probes that have sequence similarities with the mtDNA. Probes were considered to show strong evidence of cross-hybridization with the mtDNA if the probes' sequences had 90% identity over the aligned region, at least 40 of 50 matching bps, and no gaps.

4.6.2 Association analyses on DNA methylation (II)

Discovery phase

Methylation, mtDNA and clinical data were available for 926 YFS participants, and 241 variants had a VAF of ≥ 0.01. In order to reduce the computational effort,

we selected a set of 115 tagging variants that capture each of the 241 variants with $r^2 \geq 0.8$, by using Tagger (de Bakker et al. 2005) and HaploView (Barrett et al. 2005). DNAm β values were modelled as a linear function of the presence or absence of the variant allele in R. The model was adjusted for age, sex, BMI, smoking status, white blood cell type proportions, the first five PCs of DNAm array control probes, and the first six mtDNA PCs. The fraction of white blood cells (CD8T, CD4T, NK cells, B cells, monocytes, and granulocytes) was estimated through the reference-based Houseman method using the minfi package (Houseman et al. 2012). CpG loci were considered differentially methylated if they reached a Bonferroni-corrected p-value of 7.8×10^{-10} ($9 \times 10^{-8}/115$), based on the number of independent tests in the MethylationEPIC array (Mansell et al. 2019) and the number of tagging mtSNPs.

A similar linear model was applied to flag CpG sites for those showing differential methylation levels between the nine major haplogroups in 863 individuals. We selected the most common haplogroup, H, to be the reference to which other haplogroups were compared. MtDNA PCs were removed from the covariates. Significance was defined as $p < 1.0 \times 10^{-8}$ ($9 \times 10^{-8}/9$).

Sex-specific effects were studied by applying the linear model to males and females separately, and by calculating the gender-heterogeneity p-value from fixed-effect meta-analysis in the GWAMA software (Mägi et al. 2010; Mägi and Morris 2010). We required a minimum variant allele count of 10 in both sexes, which resulted in 63 tagging variants to be included. Significance was defined as $p < 1.4 \times 10^{-9}$ ($9 \times 10^{-8}/63$).

The effect of prediabetes was examined similarly by applying the model to individuals with prediabetes and normoglycaemic controls. Forty-seven tagging variants were studied, resulting in a significance threshold of 1.9×10^{-9} ($9 \times 10^{-8}/47$).

In epigenome-wide association studies (EWASs), the outcome of interest is generally associated with many small genetic effects. This results in λ_G overestimating the true inflation, as it is dependent on the number of significant associations (Yang et al. 2011). Another prominent feature especially in EWASs is confounding, which may subject test statistics to bias that is not taken into account when using λ_G (van Iterson et al. 2017; Wang et al. 2017). Therefore, the effect estimates, their standard errors and the corresponding p-values were corrected for bias and inflation using the R package bacon. The inflation function of the same package was applied to estimate the Bayesian inflation factor λ_B for each of the analyses. In EWASs, inflation

has shown to be minimal when $\lambda_B < 1.14$. (van Iterson et al. 2017.)

Replication and meta-analysis

Replication was sought in the LURIC study by applying the same linear models as in the discovery phase for 1,456 and 2,290 samples from the HumanExome-12 and 200k MetaboChip arrays, respectively. Most of the genotyped individuals ($n = 1,429$) were present in both arrays, and the total number of individuals was 2,317. Associations were considered fully replicated if the replication p-value fell below a Bonferroni-corrected p-value of $0.05/n$, with n being the number of significant associations in the discovery study covered in the replication sample. For nominal replication, the p-value threshold was set at 0.05. Consistent effect directions across both cohorts were also required.

Of the variants genotyped with the two microarrays, HumanExome-12 and 200k MetaboChip, 53 and 42 mtSNPs with a VAF of ≥ 0.005, respectively, were included. The two arrays were analysed separately, including the twelve overlapping mtSNPs present in both arrays, thus providing the opportunity of validation in the case of significant results. If a tagging mtSNP from the discovery sample was not genotyped in the replication sample, an mtSNP for replication was searched from the tagged mtSNPs. If several tagged mtSNPs were genotyped, linear regression was performed on all tagged variants and the sentinel mtSNP with the smallest association p-value was used.

The observed replication rates were benchmarked for general mtSNP and haplogroup analyses by calculating the expected degree of replication. First, we used a false-discovery rate inverse quantile transformation to correct the effect sizes for the winner's curse, an ascertainment bias where the true genetic effect of a variant is smaller than its estimate within a discovery cohort due to chance noise (Okbay et al. 2016). We also took into account the lower number of mtSNPs available in the replication cohort. Second, we calculated the expected number of associations meeting the Bonferroni-corrected replication threshold by using the method described in Okbay et al. 2016.

Finally, we performed a fixed-effect inverse-variance-weighted meta-analysis of the replicated associations by combining the effect estimates and standard errors from the discovery and replication cohorts with the GWAMA software. An association was considered significant if the meta-analysis p-value fell below the significance

threshold used in the corresponding discovery analysis.

Expression quantitative trait methylation analysis

To gain insight into whether the identified methylation QTL were connected to biological processes, we examined the associations between peripheral blood genome-wide transcriptomics and the differentially methylated CpG sites in 1,364 YFS participants. CpGs were regressed against cell count proportions and the first 30 PCs of the array control probes. Similarly, the 19,637 transcription probes were regressed against the first 20 PCs derived from the expression data. For each CpG site, expression probes within a 2 Mb window ($\pm$ 1 Mb) were included. Linear regression was applied between the residuals from the CpG regression (explanatory variable) and the expression probe residuals (dependent variable). The model was additionally adjusted for age, sex and BMI. The p-value for statistical significance was defined as 0.05, divided by the number of combinations between CpGs and genes.

4.6.3 Association analyses on blood pressure (III)

Genotype and phenotype data were available for 1,150 YFS participants. In FINCAVAS, both the genotype and phenotype data were available for 2,193 and 923 samples from the Cardio-MetaboChip and HumanCoreExome arrays, respectively. The total number of FINCAVAS individuals was 3,112 since four samples were genotyped with both arrays. For these four samples, the genotypes for the 34 overlapping variants present in both arrays were set as missing in the Cardio-MetaboChip array.

A rescaled inverse normal transformation was applied to the BP levels. This makes the distributions normal and controls the type I error, restores the original scale of measurement, deals with phenotypic outliers, and thus enhances the power of meta-analysis (Auer et al. 2016; Tang and Lin 2015). SBP, DBP and MAP were modelled as a linear function of the presence or absence of the variant allele with age, sex, BMI and a cohort-specific number of mtDNA PCs as covariates under the R environment.

A random-effect meta-analysis implemented in GWAMA was used to combine the linear regression results from both cohorts, including 87 variants with a VAF of ≥ 0.01. Using Matrix Spectral Decomposition and the eigenvalues of a variant correlation matrix (Li and Ji 2005; Nyholt 2004), we determined that 45 of the 87 variants

represented an estimate of the number of independent genetic effects for mtDNA. This resulted in a Bonferroni-corrected significance level of 0.001 (i.e. 0.05/45). We did not account for the testing of three BP traits, since they were correlated (Pearson's r was 0.78–0.96 in the YFS and 0.67–0.92 in the FINCAVAS).

Sex-specific effects were tested by applying the linear model to males and females from each cohort separately, combining the results and calculating the gender-heterogeneity p-value in GWAMA. Sixty-six variants with a VAF of ≥ 0.01 were analysed, 33 of which represented an estimate of the number of independent genetic effects. To account for the two sexes tested, the significance level was defined as $p < 7.6 \times 10^{-4}$ (0.05/33/2).

Association test of rare mtDNA variants

Standard methods for testing common variant associations are underpowered for detecting associations with rare (VAF < 0.01) variants (Li and Leal 2008). We applied a sequence kernel association test (SKAT) that collapses and tests the collective effects of variants within a specified genetic region without assuming similar directionality or effect size for each variant (Wu et al. 2011). The variants were clustered into seven regions: the four OXPHOS complexes, all rRNAs combined, all tRNAs combined, and the non-coding control region.

We employed SKAT meta-analysis implemented in the seqMeta R package with two VAF cut-off values, ≤ 0.01 (T1) and ≤ 0.05 (T5), with the default beta weights and with age, sex, BMI and cohort-specific number of mtDNA PCs as covariates. Bonferroni-corrected statistical significance was defined as $0.05/7 = 0.007$.

The pathogenicity of the nonsynonymous variants was predicted with MutPred (Mort et al. 2010; Pereira et al. 2011) and MitoTIP pathogenicity scores (Sonney et al. 2017). An additional SKAT meta-analysis was leveraged, similarly to what has been described above but including only variants with a MutPred score, > 0.5 (i.e., potentially or high-confidence harmful) or a MitoTIP classification 'possibly' or 'likely pathogenic'. Currently, no tools exist for annotating pathogenicity for variants in the rRNA regions.

5 RESULTS

5.1 Associations of mtDNA variants with the genome-wide transcriptome (I)

The characteristics of the study population used in study I are shown in Table 5.1.

Table 5.1 Characteristics of the study population in study I. Values are mean (SD) or n (%) for continuous and categorical variables, respectively.

	All	Men	Women	Prediabetes	Controls
No. of subjects	955	407	548	249	584
Age, years	42.1 (5.0)	42.1 (4.6)	42.1 (5.0)	42.8 (5.1)	41.6 (5.0)
Women	548 (57.3)	-	-	106 (42.6)	382 (65.4)
Body mass index, kg/m^2	26.7 (5.0)	27.5 (4.6)	26.1 (5.2)	28.6 (5.4)	25.5 (4.2)

In all, 3,907,763 probe-mtSNP associations were tested using linear regression. The genomic inflation factor was 1.00, indicating no inflation of the results. As shown in Table 5.2, a total of 53 expression probe-mtSNP pairs were significant after Bonferroni correction, corresponding to 5 nuclear and 2 mitochondrial genes and 31 mtSNPs. These seven genes included signal peptidase complex subunit 2 pseudogene 4 (*SPCS2P4*), ring finger protein 113A (*RNF113A*), signal peptidase complex subunit 2 (*SPCS2*), mitochondrially encoded cytochrome c oxidase II (*MT-CO2*), cardiolipin synthase 1 (*CRLS1*), solute carrier family 25 member 15 (*SLC25A14*) and mitochondrially encoded 16S RNA-like 1 (*MT-RNR2L1*).

Five associations survived all conditional analyses, taking into account the LD between variants. The two mtSNPs that associated with *MT-CO2* were in very weak LD, and only one mtSNP associated with *MT-RNR2L1*; these associations were not subjected to conditional analysis and were considered independent. For *SLC25A14* and *CRLS1*, none of the associations remained independent.

Table 5.2 The 53 significant mtSNP associations with gene transcripts. An asterisk (*) marks the associations that were replicated from the results by Kassam et al. (2016). The 'MT-' prefix is omitted from each mitochondrial gene name in the Locus column. Modified from study I.

mtSNP	Locus	VAF	Transcript	Chr	Beta	SE	p-value	
m.9055G>A	*ATP6*	0.04	SPCS2P4	1	−1.11	0.08	1.8×10^{-39}	
m.3480A>G	*ND1*	0.04	SPCS2P4	1	−1.09	0.09	4.8×10^{-32}	*
m.10550A>G	*ND4L*	0.04	SPCS2P4	1	−1.08	0.09	9.5×10^{-32}	*
m.14167C>T	*ND6*	0.04	SPCS2P4	1	−1.08	0.09	9.5×10^{-32}	
m.16224T>C	HV region 1	0.05	SPCS2P4	1	−0.95	0.08	1.8×10^{-30}	
m.11299T>C	*ND4*	0.04	SPCS2P4	1	−1.03	0.09	5.9×10^{-29}	
m.1189T>C	*RNR1*	0.03	SPCS2P4	1	−1.04	0.09	1.2×10^{-27}	*
m.9055G>A	*ATP6*	0.04	RNF113A	X	−1.23	0.13	2.5×10^{-21}	
m.9698T>C	*CO3*	0.05	SPCS2P4	1	−0.74	0.08	1.0×10^{-19}	*
m.14167C>T	*ND6*	0.04	RNF113A	X	−1.25	0.14	1.9×10^{-18}	
m.10550A>G	*ND4L*	0.04	RNF113A	X	−1.24	0.14	2.0×10^{-18}	
m.3480A>G	*ND1*	0.04	RNF113A	X	−1.24	0.14	2.7×10^{-18}	
m.9093A>G	*ATP6*	0.02	SPCS2P4	1	−1.03	0.12	4.7×10^{-18}	
m.11299T>C	*ND4*	0.04	RNF113A	X	−1.2	0.14	1.2×10^{-17}	
m.1189T>C	*RNR1*	0.03	RNF113A	X	−1.23	0.14	4.4×10^{-17}	
m.9903T>C	*CO3*	0.02	SPCS2P4	1	−1.06	0.12	7.2×10^{-17}	
m.14798T>C	*CYB*	0.10	SPCS2P4	1	−0.60	0.07	9.7×10^{-16}	
m.1811A>G	*RNR2*	0.08	SPCS2P4	1	−0.50	0.07	4.9×10^{-14}	
m.16224T>C	HV region 1	0.05	RNF113A	X	−0.96	0.13	8.6×10^{-14}	
m.9055G>A	*ATP6*	0.04	SPCS2	11	−0.93	0.12	9.6×10^{-14}	
m.8869A>G	*ATP6*	0.02	CRLS1	20	−0.75	0.10	5.3×10^{-13}	
m.4639T>C	*ND2*	0.02	CRLS1	20	−0.75	0.10	8.5×10^{-13}	
m.11377G>A	*ND4*	0.03	SPCS2P4	1	−0.76	0.11	2.2×10^{-12}	
m.9698T>C	*CO3*	0.05	RNF113A	X	−0.87	0.12	2.6×10^{-12}	
m.5263C>T	*ND2*	0.02	CRLS1	20	−0.75	0.11	5.3×10^{-12}	
m.9093A>G	*ATP6*	0.02	RNF113A	X	−1.23	0.18	7.4×10^{-12}	
m.8269G>A	*CO2*	0.01	MT-CO2	MT	−1.62	0.24	2.1×10^{-11}	*
m.11251A>G	*ND4*	0.12	SPCS2P4	1	1.04	0.15	2.6×10^{-11}	
m.15452C>A	*CYB*	0.12	SPCS2P4	1	1.03	0.15	5.6×10^{-11}	
m.9903T>C	*CO3*	0.02	RNF113A	X	−1.24	0.19	2.0×10^{-10}	
m.3505A>G	*ND1*	0.05	SLC25A14	X	0.77	0.12	3.3×10^{-10}	
m.1243T>C	*RNR1*	0.05	SLC25A14	X	0.76	0.12	4.5×10^{-10}	
m.14798T>C	*CYB*	0.10	RNF113A	X	−0.72	0.11	4.8×10^{-10}	
m.3480A>G	*ND1*	0.04	SPCS2	11	−0.85	0.14	4.8×10^{-10}	*
m.10550A>G	*ND4L*	0.04	SPCS2	11	−0.85	0.14	5.0×10^{-10}	*
m.16256C>T	HV region 1	0.07	MT-RNR2L1	MT	0.39	0.06	5.3×10^{-10}	
m.14167C>T	*ND6*	0.04	SPCS2	11	−0.85	0.14	5.4×10^{-10}	
m.16224T>C	HV region 1	0.05	SPCS2	11	−0.76	0.12	9.7×10^{-10}	
m.11947A>G	*ND4*	0.05	SLC25A14	X	0.75	0.12	1.0×10^{-9}	
m.8994G>A	*ATP6*	0.05	SLC25A14	X	0.74	0.12	1.1×10^{-9}	
m.4216T>C	*ND1*	0.12	RNF113A	X	1.32	0.22	1.4×10^{-9}	
m.5046G>A	*ND2*	0.05	SLC25A14	X	0.74	0.12	1.8×10^{-9}	
m.10398A>G	*ND3*	0.14	SPCS2P4	1	−0.40	0.07	1.8×10^{-9}	
m.15884G>C	*CYB*	0.05	SLC25A14	X	0.74	0.12	2.3×10^{-9}	
m.1811A>G	*RNR2*	0.08	RNF113A	X	−0.60	0.10	3.4×10^{-9}	
m.15452C>A	*CYB*	0.12	RNF113A	X	1.39	0.24	5.2×10^{-9}	
m.16162A>G	HV region 1	0.06	MT-CO2	MT	−0.66	0.11	5.5×10^{-9}	
m.11299T>C	*ND4*	0.04	SPCS2	11	−0.79	0.13	7.4×10^{-9}	
m.11251A>G	*ND4*	0.12	RNF113A	X	1.36	0.23	7.8×10^{-9}	
m.12414T>C	*ND5*	0.05	SLC25A14	X	0.69	0.12	8.0×10^{-9}	
m.1189T>C	*RNR1*	0.03	SPCS2	11	−0.82	0.14	8.3×10^{-9}	
m.5460G>A	*ND2*	0.05	SLC25A14	X	0.65	0.11	9.9×10^{-9}	
m.11377G>A	*ND4*	0.03	RNF113A	X	−0.93	0.16	1.2×10^{-8}	

Abbreviations: VAF, variant allele frequency; Chr, chromosome; SE, standard error; HV, hypervariable. The beta-coefficient represents the proportion of one SD change in normalized gene expression intensity (mean = 0)

The five independent associations are shown in Table 5.3, and Figure 5.1 illustrates the expression intensities for the top four transcripts relative to the alleles of the top independent locus.

Table 5.3 Independent mtSNP-probe association signals from the pairwise conditional analysis. Adapted from study I.

mtSNP	Transcript	$Beta_{cond}$	SE_{cond}	p_{cond}	p_{Bonf}
m.9055G>A	SPCS2P4	− 0.97 – − 1.36	0.09–0.20	1.8×10^{-11} – 4.4×10^{-29}	4.2×10^{-3}
m.11251A>G	SPCS2P4	0.97	0.30	1.4×10^{-3}	5.0×10^{-2}
m.15452C>A	SPCS2P4	1.13	0.24	4.3×10^{-6}	5.0×10^{-2}
m.9055G>A	RNF113A	− 1.18 – − 1.36	0.14–0.32	1.4×10^{-4} – 2.8×10^{-14}	4.2×10^{-3}
m.9055G>A	SPCS2	− 0.90 – − 1.39	0.21–0.31	3.7×10^{-5} – 1.4×10^{-6}	8.3×10^{-3}

Abbreviations and definitions: $Beta_{cond}$, effect size range from pairwise conditional analysis; SE_{cond}, standard error range from conditional analysis; p_{cond}, p-value range from conditional analysis; p_{Bonf}, mtSNP-specific Bonferroni-corrected p-value accounts for the number of pairwise analyses run for each mtSNP and is defined as the limit of significance.

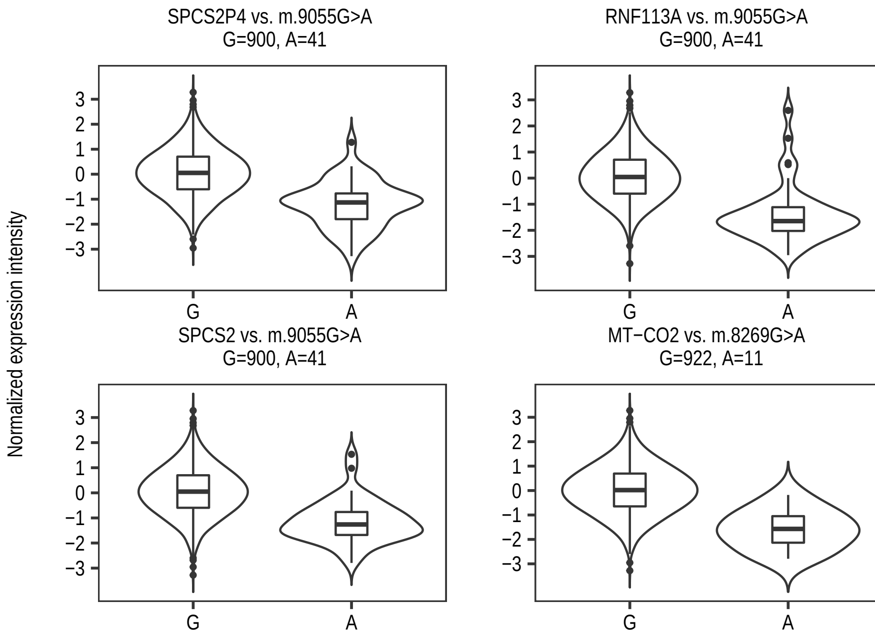

Figure 5.1 Combined box plot and violin plot of the normalized expression intensities for the top four genes relative to the alleles of the top associated independent mtSNP. Adapted from study I.

The haplogroup frequencies of the study population are shown in Table 5 in original publication I. Nearly 70% of the participants belonged to haplogroups H and U.

Five probes showed differential expression between haplogroups ($\lambda_G = 1.04$). Three of the transcripts, SPCS2P4, RNF113A and SLC25A14, were also associated with individual mtSNPs, but the other two transcripts, solute carrier family 2 member 8 (SLC2A8) and mitochondrially encoded NADH dehydrogenase 5 (MT-ND5), were not identified in the probe-mtSNP analysis. The significant associations identified by Tukey's post hoc test are shown in Table 5.4. Figure 5.2 illustrates the transcript levels of the top two genes, *SPCS2P4* and *RNF113A*, which were significantly lower in haplogroup K than in all other eight major haplogroups. This haplogroup is defined by the same variants (m.9055G>A, m.3408A>G and m.10550A>G, for example) that constitute the vast majority of the associations with the transcripts SPCS2P4 and RNF113A in the mtSNP-based analysis.

Table 5.4 Significant and independent mitochondrial genome-wide associations with gene transcripts. Adapted from study I.

Transcript	Haplogroup comparison	Difference in means [95% CI]	Tukey-adjusted p-value
SPCS2P4	K-H	−1.27 [−1.80, −0.73]	1.3×10^{-11}
	K-U	−1.23 [−1.78, −0.68]	2.4×10^{-10}
	K-J	−1.30 [−1.93, −0.67]	8.6×10^{-9}
	K-W	−1.35 [−2.04, −0.66]	6.1×10^{-8}
	K-V	−1.17 [−1.82, −0.52]	9.9×10^{-7}
	K-T	−1.10 [−1.78, −0.43]	1.6×10^{-5}
	K-I	−1.47 [−2.38, −0.55]	2.7×10^{-5}
	K-X	−1.28 [−2.14, −0.43]	1.1×10^{-4}
RNF113A	K-H	−1.44 [−1.97, −0.91]	$<2.6\times10^{-14}$
	K-U	−1.53 [−2.07, −0.99]	$<2.6\times10^{-14}$
	K-J	−1.52 [−2.14, −0.90]	2.5×10^{-12}
	K-T	−1.44 [−2.10, −0.77]	1.1×10^{-9}
	K-W	−1.42 [−2.10, −0.74]	4.7×10^{-9}
	K-X	−1.48 [−2.32, −0.65]	1.7×10^{-6}
	K-I	−1.57 [−2.47, −0.67]	2.8×10^{-6}
	K-V	−1.03 [−1.67, −0.39]	2.2×10^{-5}
	K-U	−0.50 [−0.94, −0.06]	1.2×10^{-2}
SLC25A14	W-J	0.81 [0.21, 1.41]	1.0×10^{-3}
	W-H	0.55 [0.06, 1.05]	1.6×10^{-2}
	K-J	0.69 [0.05, 1.33]	2.5×10^{-2}
	W-X	0.86 [0.02, 1.69]	3.8×10^{-2}
SLC2A8	U-J	−0.46 [−0.88, −0.04]	2.1×10^{-2}
	U-H	−0.25 [−0.50, 0.00]	4.3×10^{-2}
MT-ND5	J-V	0.84 [0.30, 1.41]	7.1×10^{-5}
	J-U	0.59 [0.16, 1.05]	5.8×10^{-4}
	J-H	0.50 [0.10, 1.33]	3.2×10^{-3}
	J-W	0.69 [0.09, 1.29]	1.1×10^{-2}

Abbreviations: CI, confidence internal. One unit of difference in means represents the proportion of one SD change in normalized gene expression intensity.

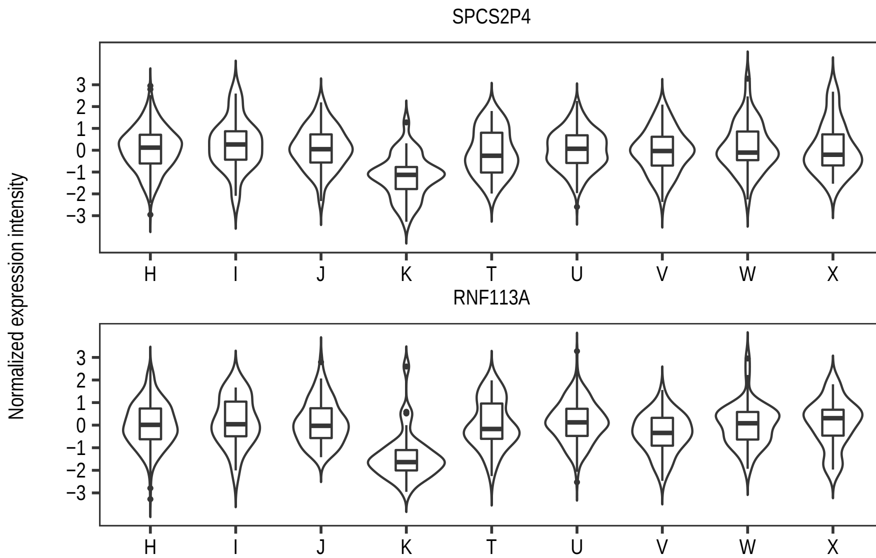

Figure 5.2 Combined box plot and violin plot of the normalized expression intensities of *SPCS2P4* and *RNF113A* across the nine major haplogroups. Adapted from study I.

The nucleous-encoded transcripts identified to be associated with mtSNPs or haplogroups were SPCS2P4, RNF113A, SPCS2, CRLS1, SLC25A14, and SLC2A8. None of the corresponding genes showed cross-hybridization with sequences on the mitochondrial genome.

5.2 Associations of mtDNA variants with DNA methylation (II)

The characteristics of the study populations used in study II are shown in Table 1 in original publication II. The LURIC study participants were, on average, older than the YFS participants, with a higher percentage of men and individuals with prediabetes. The proportion of current smokers was similar in both cohorts, but the percentage of never-smokers was higher in the YFS.

A total of 88,513,545 CpG–mtSNP pairs were tested in the discovery phase, resulting in 5,562 significant associations that constituted of 89 mtSNPs and 4,618 CpG sites scattered all around the nuclear genome. The Bayesian inflation factor was 1.00, suggesting minimal inflation. Of the 5,562 associations, 685 were available for replication, resulting in a significance level of $p < 7.3 \times 10^{-5}$ (0.05/685). None of the as-

sociations passed this threshold, even though 228 associations were expected to reach this p-value. Twenty-one associations were nominally replicated with $p < 0.05$. The fixed-effect meta-analysis that combined the nominally replicated results yielded 19 associations with epigenome-wide significance (7.8×10^{-10}), shown in Figure 5.3 and in Table 5.5. The significant CpG–mtSNP pairs did not correspond to the transcript-mtSNP pairs that were identified in study I.

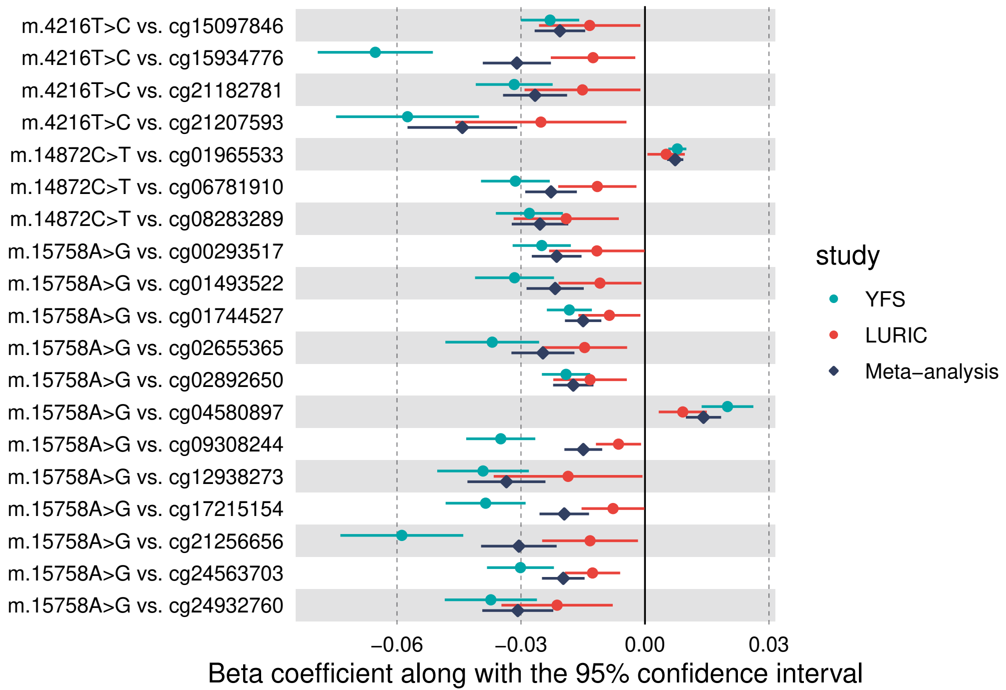

Figure 5.3 Forest plot showing the 19 nominally replicated mtSNP effects on DNAm that also reached epigenome-wide significance in the meta-analysis. Adapted from study II.

Table 5.5 Nineteen nominally replicated CpG–mtSNP associations that reached epigenome-wide significance in the meta-analysis. For all LURIC associations, the corresponding mtSNP was genotyped only with the HumanExome-12 array. Modified from study II.

				YFS			LURIC			Combined		
mtSNP	VAF	CpG	Reference gene	Beta	SE	p	Beta	SE	p	Beta	SE	p
m.4216T>C[1]	0.12	cg15934776	*AFF3*	−0.065	0.007	$4.4{\times}10^{-20}$	−0.013	0.005	$1.6{\times}10^{-2}$	−0.031	0.004	$1.8{\times}10^{-13}$
m.4216T>C[1]	0.12	cg21182781	*IARS2*	−0.032	0.005	$2.6{\times}10^{-11}$	−0.015	0.007	$3.5{\times}10^{-2}$	−0.027	0.004	$1.8{\times}10^{-11}$
m.4216T>C[1]	0.12	cg21207593	*LIG3*	−0.057	0.009	$7.5{\times}10^{-11}$	−0.025	0.011	$1.7{\times}10^{-2}$	−0.044	0.007	$7.1{\times}10^{-11}$
m.4216T>C[1]	0.12	cg15097846	*TULP3*	−0.023	0.004	$1.8{\times}10^{-10}$	−0.013	0.006	$3.3{\times}10^{-2}$	−0.021	0.003	$4.4{\times}10^{-11}$
m.14872C>T[2]	0.020	cg06781910	-	−0.030	0.004	$1.5{\times}10^{-13}$	−0.010	0.005	$1.7{\times}10^{-2}$	−0.023	0.003	$1.1{\times}10^{-12}$
m.14872C>T[2]	0.020	cg01965533	*DLST*	0.008	0.001	$2.6{\times}10^{-12}$	0.005	0.002	$2.7{\times}10^{-2}$	0.007	0.001	$3.9{\times}10^{-13}$
m.14872C>T[2]	0.020	cg08283289	*ZC3H13*	−0.028	0.004	$1.4{\times}10^{-11}$	−0.020	0.006	$3.4{\times}10^{-3}$	−0.025	0.003	$3.5{\times}10^{-13}$
m.15758A>G	0.014	cg09308244	-	−0.035	0.004	$2.9{\times}10^{-16}$	−0.006	0.003	$2.0{\times}10^{-2}$	−0.015	0.002	$1.6{\times}10^{-10}$
m.15758A>G	0.014	cg17215154	*IQCE*	−0.039	0.005	$5.7{\times}10^{-15}$	−0.008	0.004	$4.7{\times}10^{-2}$	−0.020	0.003	$1.7{\times}10^{-10}$
m.15758A>G	0.014	cg21256656	*KLK6*	−0.059	0.008	$8.8{\times}10^{-15}$	−0.013	0.006	$2.5{\times}10^{-2}$	−0.031	0.005	$6.3{\times}10^{-11}$
m.15758A>G	0.014	cg24563703	-	−0.030	0.004	$3.0{\times}10^{-13}$	−0.013	0.003	$2.1{\times}10^{-4}$	−0.020	0.002	$6.1{\times}10^{-14}$
m.15758A>G	0.014	cg00293517	-	−0.025	0.004	$3.6{\times}10^{-12}$	−0.012	0.006	$4.9{\times}10^{-2}$	−0.021	0.003	$3.5{\times}10^{-12}$
m.15758A>G	0.014	cg12938273	*GNAZ*; *RSPH14*	−0.039	0.006	$4.3{\times}10^{-12}$	−0.019	0.009	$4.3{\times}10^{-2}$	−0.034	0.005	$3.5{\times}10^{-12}$
m.15758A>G	0.014	cg01744527	*SCAF1*	−0.018	0.003	$4.4{\times}10^{-11}$	−0.009	0.004	$2.5{\times}10^{-2}$	−0.015	0.002	$2.9{\times}10^{-11}$
m.15758A>G	0.014	cg24932760	*C17orf67*	−0.037	0.006	$5.6{\times}10^{-11}$	−0.021	0.007	$2.0{\times}10^{-3}$	−0.031	0.004	$2.3{\times}10^{-12}$
m.15758A>G	0.014	cg01493522	-	−0.032	0.005	$9.5{\times}10^{-11}$	−0.011	0.005	$3.4{\times}10^{-2}$	−0.022	0.004	$7.7{\times}10^{-10}$
m.15758A>G	0.014	cg02892650	*LRRC36*	−0.019	0.003	$1.5{\times}10^{-10}$	−0.013	0.005	$3.5{\times}10^{-3}$	−0.017	0.003	$3.5{\times}10^{-12}$
m.15758A>G	0.014	cg02655365	*STK38*	−0.037	0.006	$1.7{\times}10^{-10}$	−0.015	0.005	$5.5{\times}10^{-3}$	−0.025	0.004	$2.2{\times}10^{-10}$
m.15758A>G	0.014	cg04580897	*SNORA6*; *RPSA*	0.020	0.003	$4.1{\times}10^{-10}$	0.009	0.003	$2.1{\times}10^{-3}$	0.014	0.001	$7.7{\times}10^{-11}$

Abbreviations: VAF, variant allele frequency in the discovery sample; SE, standard error; [1] tagged mtSNP m.15452C>A used in LURIC; [2] tagged mtSNP m.2259C>T used in LURIC.

The haplogroup-based discovery analysis identified 142 significant associations with minimal inflation ($\lambda_B = 0.99$). In the LURIC study, 120 associations were available for replication, none of which were fully replicated with $p < 4.2 \times 10^{-4}$ (0.05/120). Fifteen associations were expected to reach this level. Six associations were nominally replicated, four of which remained significant in the meta-analysis (Table 5.6 and Figure 5.4). Similarly to the mtSNP-based analysis, the haplogroup associations did not correspond to the haplogroup–transcript associations that were identified in study I.

Table 5.6 Four nominally replicated CpG–haplogroup associations that reached epigenome-wide significance in the meta-analysis. Modified from study II.

			YFS		LURIC		Combined	
Haplo-group	CpG	Ref. gene	Beta (SE)	p-value	Beta (SE)	p-value	Beta (SE)	p-value
W	cg25821304	*RNF135*	0.005 (0.001)	6.1×10^{-10}	0.006 (0.001)	8.2×10^{-4}	0.005 (0.001)	2.2×10^{-12}
I	cg20934571	*CARKD*	−0.021 (0.003)	1.3×10^{-9}	−0.018 (0.006)	1.0×10^{-3}	−0.020 (0.003)	5.9×10^{-12}
I	cg11350158	*LRP1B*	−0.012 (0.002)	1.7×10^{-9}	−0.005 (0.002)	3.3×10^{-2}	−0.009 (0.001)	2.6×10^{-9}
I	cg25020969	*MAD1L1*	−0.034 (0.006)	9.8×10^{-9}	−0.014 (0.006)	1.3×10^{-2}	−0.024 (0.004)	9.3×10^{-9}

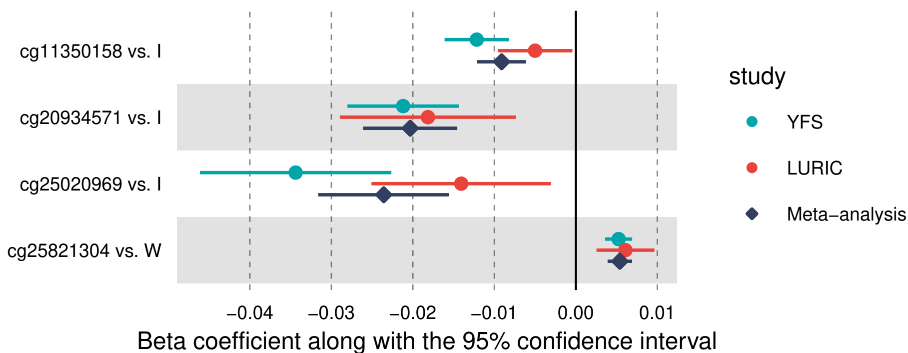

Figure 5.4 Forest plot showing the four nominally replicated haplogroup effects on DNAm that also reached epigenome-wide significance in the meta-analysis. In all associations, haplogroup H was used as the reference haplogroup. Adapted from study II.

Finally, we searched for gene transcripts regulated by the candidate CpGs identified after the replication stage. We considered genes ±1 Mb from each CpG site and

tested 890 CpG–transcript combinations for differential expression. Two associations were significant after correction for multiple testing: inverse associations were observed for cg25020969 (which showed lower methylation levels in haplogroup I) and probes ILMN_1681674 and ILMN_2358069, both at the *MAD1L1* gene (effect estimate –4.63 and –3.68, SE 1.05 and 0.91, p-value 1.2×10^{-5} and 5.9×10^{-5}, respectively).

5.3 Prediabetes-specific effects (I and II)

In study I, the meta-analysis showed a difference in the effect sizes between subjects with prediabetes and controls for one transcript. A heterogeneity p-value of 8.9×10^{-9} ($\lambda_G = 0.99$) corresponded to the association between the transcript levels of 2'-5'-oligoadenylate synthase like (*OASL*) and m.16294C>T. For this mtSNP, individuals with prediabetes had an effect estimate of –0.74 (SE 0.12) and a corresponding p-value of 9.7×10^{-6} ($\lambda_G = 0.99$), while the control group had an effect estimate of 0.43 (SE 0.12) and a p-value of 4.2×10^{-4} ($\lambda_G = 1.00$). In other words, individuals with prediabetes and the allele T had lower transcript levels of *OASL* compared to the reference allele C, while participants with the T allele but no prediabetes had higher transcript levels compared to the reference allele, as can be seen in Figure 5.5.

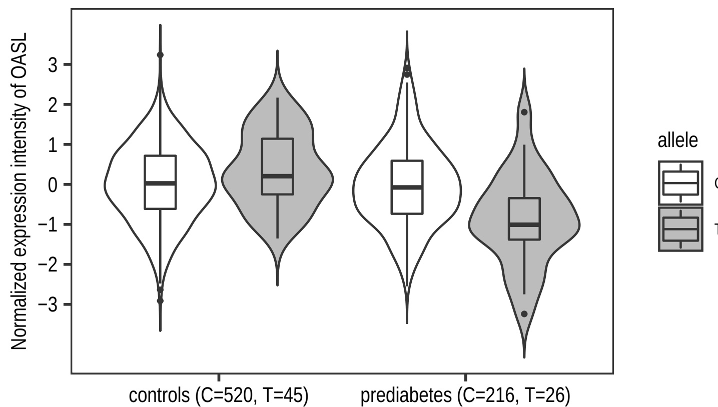

Figure 5.5 Combined boxplot and violin plot of the normalized expression intensities of *OASL* across the individuals with and without prediabetes relative to mtSNP m.16294C>T. Adapted from study I.

In study II, prediabetes-specific analysis discovered 483 mtSNP–CpG pairs that

showed significant differences in effect sizes between the phenotypes. We were able to attempt replication for 113 associations in the LURIC study, none of which were replicated with a heterogeneity p-value of < 0.05.

5.4 Mitochondrial GWAS on blood pressure (III)

The characteristics of the study populations used in study III are shown in Table 1 in original publication III. The participants in the FINCAVAS were, on average, older, had higher BMI and BP levels and displayed clearly more antihypertensive medication usage (31%, 24% and 16% for 1, 2 and ≥ 3 medication classes taken, respectively) than YFS participants (7%, 3% and 0.3%, respectively).

The associations of 87 common mtSNPs with SBP, DBP and MAP were evaluated using random-effect meta-analysis in these two cohorts. When both sexes were analysed together, no statistically significant associations were observed after correction for multiple testing. Eight associations were nominally significant, with $p < 0.05$, and are shown in Table 5.7.

Table 5.7 The eight nominally significant associations with BP. Adapted from study III.

mtSNP	Locus	VAF	Trait	Beta	SE	p-value	n
m.1243C>T	*MT-RNR1*	0.02	SBP	−4.2	1.8	0.019	4,219
m.15257A>G	*MT-CYB*	0.01	MAP	4.9	2.2	0.025	2,071
m.11674T>C	*MT-ND4*	0.03	SBP	−4.3	1.9	0.026	3,342
m.4024G>A	*MT-ND1*	0.02	DBP	4.2	2.0	0.033	2,040
m.5004C>T	*MT-ND2*	0.02	DBP	4.2	2.0	0.036	2,035
m.4336C>T	tRNA	0.01	MAP	−4.1	2.0	0.037	3,310
m.9055G>A	*MT-ATP6*	0.05	DBP	2.6	1.3	0.042	3,334
m.4336C>T	tRNA	0.01	DBP	−3.5	1.8	0.043	3,310

Finally, we conducted SKAT meta-analyses on all rare (T1 test) and low-frequency (T5 test) variants. We also employed SKAT taking into account the predicted pathogenicity of the variants. None of the analyses yielded significant associations with the BP traits over the tested mtDNA regions.

5.5 Sex-specific effects (I–III)

In study I, the meta-analysis did not show significant heterogeneity in the mtSNP–transcript effect sizes between the sexes. In the discovery stage in study II, sex analysis discovered 664 mtSNP–CpG pairs that showed significant differences in the effect sizes between males and females. Replication was attempted for 135 associations in the LURIC study, none of which were replicated with a heterogeneity p-value of < 0.05. Finally, in study III, applying the BP association analysis to males and females separately and calculating the gender-heterogeneity p-value did not identify significant mtSNP–BP associations showing sexual dimorphism.

6 DISCUSSION

6.1 Transcriptomic changes associated with mtDNA variation (I)

The mitochondrial GWAS on gene expression replicated seven associations that have been previously reported (Kassam et al. 2016) and identified 46 novel mtSNP–probe associations. The mtDNA sequence used by Kassam et al. is based on the Yoruban sequence, and the mtDNA position values need to be converted to match the rCRS: the corresponding rCRS position is obtained by decreasing the Yoruban position by one for Yoruban base pair range 3109–16190, and by two for Yoruban range 318–3108 (Lott et al. 2013). For example, variants m.3481A>G and m.1191T>C in the Yoruban sequence correspond to rCRS variants m.3480A>G and m.1189T>C, respectively.

Most of the associations were located in the protein-encoding genes and few in the HV region and rRNA-encoding genes. No associations were identified for mtSNPs in the tRNA-encoding genes. The replicated associations correspond to both nuclearly and mitochondrially encoded genes (*SPCS2P4*, *SPCS2* and *MT-CO2*) and strengthen the hypothesis that mitochondria have a genetic control of the expression of several genes. *SPCS2P4* is a pseudogene, the protein product of *SPCS2* is involved in the mitotic cell cycle pathway, and *MT-CO2* encodes a subunit of OXPHOS complex IV. Even though the biological relevance of these associations remains unclear and they do not necessarily result from causal relationships, they may equally represent retrograde signals that induce changes in nuclear gene expression in order to regulate mitochondrial functionality.

A potential candidate target for retrograde signalling is *SLC25A14*, encoding mitochondrial uncoupling protein 5, which was associated with several mtSNPs before conditional analysis and with haplogroup W. There is evidence that mitochon-

drial uncoupling protein 5 overexpression can modify the mitochondrial membrane potential, maintain OXPHOS and decrease the mitochondrial production of ROS (Kim-Han et al. 2001; Kwok et al. 2010). Taken together, it could be suggested that variants defining haplogroup W result in minor changes in OXPHOS that are compensated with retrograde signals.

The main finding in the haplogroup-based analysis was that haplogroup K carriers have significantly lower transcript levels of *SPCS2P4* and *RNF113A* compared to K non-carriers. This is not surprising, since most of the top mtSNPs associated with these expression probes are the defining variants for haplogroup K (van Oven 2015). Haplogroup K has been associated with an increased risk of breast cancer (Bai et al. 2007), although this association is controversial (Salas et al. 2014). Moreover, an increased RNF113A plasma protein level has been suggested as an early detection biomarker for triple-negative breast cancer (Li et al. 2012). In our study, however, haplogroup K was associated with lower transcript levels of *RNF113A*. Even though the mRNA levels may not be sufficient for predicting the protein levels, our finding suggests that, if this biomarker were adopted for clinical use, it might not be applicable to haplogroup K carriers.

6.2 DNA methylation changes associated with mtDNA variation (II)

In the discovery stage of the general mtSNP-based analysis, we observed numerous significant candidates for methylation QTL mtSNPs. Twenty-one associations achieved nominal replication, 19 of which remained epigenome-wide-significant when the discovery and replication results were combined by a fixed-effect meta-analysis. Haplogroup-based analysis pinpointed six nominally replicated associations, four of which remained significant after meta-analysis.

By reaching borderline significance in the replication analysis and epigenome-wide significance in the meta-analysis, the most promising association corresponded to haplogroup W and cg25821304 mapping to the gene *RNF135*. The protein coded by *RNF135* is involved in the innate immunity against RNA virus infections (Oshiumi et al. 2009), and mtDNA can activate specific innate immunity responses (Kenney et al. 2014a; West and Shadel 2017). The current finding might be regarded as

additional evidence that mtDNA variation could have immunomodulatory features, even though the CpG site did not associate with mRNA transcripts.

We did not observe differential methylation at CpG sites mapping to the genes that showed mtSNP- or haplogroup-specific transcriptome profiles in study I or in the work by Kassam et al. 2016. In addition, the haplogroup-K-defining variants, that constituted the majority of the transcription QTL mtSNPs in study I, were not significantly associated with DNAm levels at any stage in study II. This implies that, if there is a causal relationship between the haplogroup-K-defining variants and peripheral blood transcriptomics, the transcriptional changes are not mediated by the methylation of CpG sites.

An interesting association was observed between mtSNP m.14872C>T, a synonymous mutation in the *MT-CYB* gene, and cg01965533, which maps to dihydrolipoamide S-succinyltransferase (*DLST*). The protein product of *MT-CYB* is a subunit of OXPHOS complex III (Vázquez-Acevedo et al. 1993), and *DLST* encodes a subunit of the α-KG dehydrogenase complex which catalyses the conversion of α-KG to succinyl-CoA (Tretter and Adam-Vizi 2005). In regulatory T cells, the loss of complex III increases DNAm and leads to an accumulation of the metabolites 2-HG and succinate that inhibit the TET demethylases (Weinberg et al. 2019). Therefore, the slight increase in the methylation of cg01965533 could compromise the increase in the α-KG-derived metabolites by repressing the expression of *DLST* and thus the conversion of α-KG.

Variant m.4216T>C, a nonsynonymous mutation in the *MT-ND1* gene, was associated with lower methylation levels at four CpG sites. Two of those CpGs map to a gene that encodes a protein targeted in the mitochondria: *IARS2* and *LIG3*. The former encodes an isoleucyl-tRNA synthetase that catalyses the amino acid attachment into a cognate mt-tRNA (Vona et al. 2018). One isoform of the protein encoded by *LIG3* is the only DNA ligase in mitochondria, and it is involved in mtDNA replication and repair (Lakshmipathy and Campbell 1999). The overexpression of *LIG3* in mitochondria also protects cells against oxidative stress (Akbari et al. 2014). Variant m.4216T>C, together with the three variants tagged by it, define haplogroup JT (van Oven and Kayser 2009). The polymorphism affects the assembly or stability of OXPHOS complex I (Hinttala et al. 2010), a major site for mitochondrial ROS production (Nissanka and Moraes 2018). Sub-haplogroup J has showed lower ROS production compared to haplogroup H (Bellizzi et al. 2012; Kenney et al. 2013).

Taken together, altered methylation at cg21207593 mapping to *LIG3* could result from lower ROS levels associated with m.4216T>C.

However, in the meta-analysis combining the replicated results, the CpGs that were associated with mtSNPs were not the same as those that were associated with the haplogroups, and the mtSNPs (and the tagged polymorphisms) are not the defining variants for haplogroups W and I pinpointed in the haplogroup-based analysis (van Oven and Kayser 2009). This could imply that the haplogroup effect is driven by small individual effects by many mtSNPs. Haplogroup JT, defined by m.4216T>C and three other mtSNPs, further divides into haplogroups J and T that have both shared and unique defining variants that were not associated with CpGs. We performed association analyses separately for haplogroups J and T but not for JT, which could explain why the four CpGs were not identified in the haplogroup-based analysis.

Even though we used a relatively liberal replication threshold, the observed replication rates were significantly lower than predicted. The lack of replication was not explained by the winner's curse. The poor replication rates may be explained by study-specific heterogeneity; the YFS is a population-based study, whereas the LURIC participant pool mainly consists of older patients referred to coronary angiography. DNAm variation has been associated with age numerous times (Horvath 2013; Zhang et al. 2020), and other confounding factors, such as socioeconomic status (Fiorito et al. 2019) and lipid composition (Gomez-Alonso et al. 2021), may also have had an independent effect on the results. In fact, inconsistency or study-specific heterogeneity of genetic association results can be observed whether the associations are true or not (Ioannidis 2007). The studies may also have been underpowered to detect small genetic effects.

Two differentially methylated CpG sites were associated with peripheral blood transcripts. These associations, both corresponding to cg25020969, were not surprising, as the target of this CpG site and the two associated transcripts were related to the same gene, *MAD1L1*. It should be noted, however, that rather than being a strict dynamic mechanism for regulating gene expression, DNAm changes can also serve as a long-term memory of previous gene expression decisions that were mediated by transcriptional factors that might no longer be present in the cell (Dor and Cedar 2018).

6.3 Prediabetes-specific effects (I and II)

The association between transcript OASL and mtSNP m.16294C>T showed prediabetes-specific effects. In a small study of Bangladeshi individuals, this variant demonstrated a protective role against T2DM (Saha et al. 2021). Among the European population, on the other hand, the variant has been associated with obesity (Ebner et al. 2015) and coronary artery disease (Mueller et al. 2011). The variant is located in the control region and is possibly involved in accelerating mtDNA synthesis to satisfy developmental, physiological or ageing-related demands (Fish et al. 2004). Functionally, *OASL* encodes an interferon response protein, and its high expression levels in visceral adipose tissue, together with other three interferon signature genes, have been found to be positively correlated with adipose tissue and systemic insulin resistance (Ghosh et al. 2016). Regarding T2DM, the same gene was dysregulated in pancreatic tissue at the mRNA level when compared to healthy control individuals (Pedersen et al. 2021). The *OASL* gene has shown effects on C-reactive protein, γ-glutamyl transferase and LDL cholesterol, all of which are related to cardiovascular diseases, which are prevalent in individuals with T2DM (Middelberg et al. 2011). Based on this, it could be hypothesized that m.16294C>T has a minor role in the development of T2DM-related cardiovascular disease. However, the identified association should be replicated in other populations and explored in functional genomic studies to gain support for this hypothesis.

Variant m.16294C>T was also analysed in study II, and in the discovery phase it showed prediabetes-specific associations for seven CpGs, none of which mapping to *OASL*. These seven associations were not replicated in the LURIC study. In general, replication analysis demonstrated no evidence of prediabetes-specific mitochondrial genetic control of DNAm. As individuals with prediabetes may oscillate between normo- and dysglycaemia (Tabák et al. 2012), there is, however, a possibility that the cumulative glycaemic burden is not high enough in order to yield detectable changes at a methylomic or transcriptomic level.

6.4 Role of mtDNA variation in blood pressure (III)

The mitochondrial GWAS on BP did not identify any significant associations after correction for multiple testing, also failing to replicate the results where mtSNPs

m.3197T>C and m.15924A>G were associated with higher SBP and MAP, respectively, in white North American individuals (Buford et al. 2018). The pooled effects on SBP across all tRNA regions, also reported by Buford et al., were also not replicated, not even when the predicted pathogenicity of the variants within protein-encoding and the tRNA regions was accounted for. However, the participants in the aforementioned study were significantly older than the Finns in our cohorts, and the possible regulative role may be activated only in later life. Another variant that has been associated with BP is m.5913G>A that was identified in the Framingham Heart Study (Liu et al. 2012). This mtSNP was rare in the YFS population (VAF < 0.01) and not included in the two genotyping arrays used in the FINCAVAS, thus preventing the attempt at replication.

Our inability to replicate the findings from well-defined cohorts underlines the need of larger multi-centre consortium studies. For example, the CHARGEmtDNA+ working group analysed the associations of common and rare mtDNA variants with seven metabolic traits (not including BP) in ~170,000 individuals (Kraja et al. 2019). However, the role of mtDNA variation in BP seems to be small compared to nDNA polymorphisms, and we should also be ready to accept the null hypothesis already implied by several preceding studies (Hudson et al. 2014; Saxena et al. 2006; Venter et al. 2017). Another hypothesis to be tested is that, instead of being causal to BP variation, mtSNPs could impact the complications of hypertension, similarly to T2DM (Achilli et al. 2011).

6.5 No evidence of sexually dimorphic associations (I–III)

Similarly to the findings of Kassam et al. (2016), we did not find evidence of sex-specific mitochondrial genetic effects on gene expression. These results are in contrast with the strong, mostly male-biased mitochondrial genetic control of nuclear gene expression observed in *Drosophila melanogaster* (Innocenti et al. 2011). Since the nDNA backgrounds can modify the sensitivity of the transcriptome to mtDNA variation in *Drosophila* (Mossman et al. 2016), it is possible that the nDNA backgrounds in the YFS population are also less sensitive to mtDNA variation.

We were also unable to identify convincing sex-specific mitochondrial genetic effects on DNAm, partly due to the relatively low number of associations for which replication could even be attempted. Furthermore, blood pressure levels were not

affected by male- or female-specific genetic effects. Taken together, there is little evidence in humans of the sex-specific mitochondrial effects observed in other species, especially in *Drosophila* (Nagarajan-Radha et al. 2020). Still, the existence of 'mother's curse' mtDNA mutations in humans may be a broad phenomenon, but their deleterious effects have been counter-adapted by nuclear compensatory modifier mutations (Beekman et al. 2014), or the nuclear counter-adaptations are not effective enough to be detected within a population (Dowling and Adrian 2019).

Peripheral blood DNAm and genome-wide transcript levels and BP may not have been the optimal traits for trying to reveal the sex-specific effects. If a trait is virtually identical in both sexes, there is little basis for mtDNA variation to affect males and females differently. Therefore, the most potential traits for identifying the 'mother's curse' effects are those that are sexually highly dimorphic but analogs in function and consequences. Examples of such traits are those related to reproductive success. (Dowling and Adrian 2019.) However, as sex-specific autosomal methylation and transcription patterns have been demonstrated in peripheral blood in humans (Jansen et al. 2014; Singmann et al. 2015), we would have expected to observe some evidence of mitochondrial sex-specific genetic effects.

6.6 Strengths and limitations

A strength of all three studies was the use of large, well-characterised cohorts. A major strength was also the use of NGS in identifying the mtDNA variants in the YFS. Compared to microarray genotyping, this allowed us to study a dense set of mtSNPs and improved the quality of haplogroup assignment. In studies II and III, the LURIC and FINCAVAS participants were genotyped using microarrays, and fully comparable mtDNA sequencing datasets were not available for these two cohorts, limiting the number of mtSNPs to be included in the replication and meta-analyses. This was the main weakness in these two studies. This was especially pronounced in study II, in which we were able to attempt replication only for 15%–30% of the mtSNP-based associations from the discovery phase.

A strength of study I was that we were able to compare our associations from the general mtSNP-based analysis with the previously described results by Kassam et al. 2016. However, we did not attempt replication for the prediabetes-specific effects, and the possibility of false-positive associations should be acknowledged. Another

limitation is that no OGTTs were performed and the definition of prediabetes was based only on FPG and HbA1c levels; this also applies to the YFS participants in study II.

In study II, genotyping a part of the LURIC cohort with two different microarrays increased the quality of haplogroup assignment. The discovery analyses were adjusted for bias and inflation using a state-of-the-art method specifically developed for EWASs. The measurement of serum cotinine levels permitted us to verify the self-reported smoking status in the LURIC study, whereas in the YFS the status was only self-reported. The use of tagging mtSNPs may have resulted in false-positive or false-negative replications, and software specifically designed for QTL analyses of large datasets, such as Matrix eQTL (Shabalin 2012), would have enabled to analyse all mtSNPs within a reasonable time.

A strength of study III was the use of a large range of BP variation in the two cohorts, and the method we used to adjust for antihypertensive treatment effects has proven to work across a wire variety of clinical scenarios (Cui et al. 2003; Tobin et al. 2005). The majority of the rare mtSNPs were identified only in the YFS population, which decreased the power of the SKAT meta-analysis.

In the FINCAVAS in study III, mtDNA variants were determined from peripheral blood leukocytes, whereas peripheral whole blood was used as the source of mtDNA in the YFS. In addition to leukocytes, peripheral blood also contains erythrocytes and platelets. The former lack both nuclei and mitochondria, while the latter are also devoid of nuclei but are equipped with mitochondria (Melchinger et al. 2019). This could induce bias in mtDNA copy number studies (Hurtado-Roca et al. 2016), but the possible effect in mtSNP association studies remains unclear. Most ($>$ 90%) homoplasmic variants that have been identified from whole blood have also been present in lymphoblastoid cell lines derived from B-lymphocytes (Liu et al. 2022), and different white blood cell types display similar mtSNP profiles (Zhang et al. 2016). Taken together, this could suggest that using both peripheral blood and leukocyte-derived mtDNA data did not result in significantly different mtSNP profiles. Platelet mtDNA methylation may, however, contribute to the development of cardiovascular diseases through the modulation of platelet activity (Baccarelli and Byun 2015).

Furthermore, the participants in the YFS, LURIC and FINCAVAS studies are almost entirely white people of European descent, and the results are not directly

generalizable to other ethnic groups with different genetic and environmental backgrounds.

All mtDNA samples were derived from blood. As metabolic profiles vary between different tissues, our results may not be generalizable to other tissue types. The possible effect of tissue-specific mtSNP profiles remains unclear. Especially tissue-specific heteroplasmy is common (Naue et al. 2015; Samuels et al. 2013), but in a small study of 35 patients, the homoplasmic mtDNA variants obtained from blood and muscle were nearly identical (Taylor et al. 2003). In addition, the blood transcriptome (Melé et al. 2015) and DNA methylome (Lowe et al. 2015) display distinctive profiles compared with other somatic tissues, which further limits the generalizability of our results to other tissues types.

This study did not address the role of heteroplasmic variants. Blood mtDNA typically exhibits less heteroplasmy than tissues with higher metabolic activity (Naue et al. 2015). In the YFS population, a heteroplasmy level between 0.05 and 0.95 in at least 1% of the samples was observed only for one mtSNP, m.16192C>T, in the hypervariable region. This is in line with previous findings (Stewart and Chinnery 2021), as the YFS participants are healthy middle-aged individuals. However, as reviewed in section 2.2.1, low-level heteroplasmy is also common in healthy individuals. With a sequencing coverage of ~500x, which was achieved in the YFS population, the 0.05 heteroplasmy detection level is quite reliable, but for lower levels, a deeper sequencing (>1000x) would have been required (González et al. 2020; Weissensteiner et al. 2021). This left us unable to study the effects of low-level heteroplasmies. Finally, all three studies examined only the nucleotide substitutions, whereas deletions and insertions were not identified.

6.7 Future perspectives

It is quite possible that the identified associations are examples of mtDNA variation affecting DNAm and gene expression to maintain the homeostasis of normal cellular function. Functional genomic studies, possibly utilizing mtDNA cybrids in which the nDNA is held constant, would allow further exploration of this hypothesis and pinpoint the molecular underpinnings behind the associations. Causal interference methods, such as Mendelian randomization, could also be used to interrogate the causal effects of the mtDNA markers. On the other hand, due to the dual genomic

origin of mitochondrial proteins and the constant interaction between mitochondria and the nucleus, future studies should also address the nuclear genetic background to gain a greater understanding of the genetics behind mitochondrial–nuclear communication. In addition, the analysis of different tissues, such as skeletal muscle or liver, could provide further molecular-level and clinical implications.

Future research would also benefit from using NGS in identifying the variants, by allowing a broader range of mtSNPs to confirm the candidate associations identified in the YFS population. Furthermore, association studies utilizing data sets from several different cohorts would improve the power to detect novel associations and the generalizability of the results. Another basis for future research could be achieved by introducing more homogeneous study groups with less confounding effects.

A twenty-four hour ambulatory BP monitoring would allow the investigation of the nocturnal BP and the circadian BP variability. One mitochondrial GWAS has utilised ambulatory BP levels (Venter et al. 2017), but the mitochondrial genetics behind the night-time BP drop remain to be examined.

Finally, we did not study the effect of T2DM on mtDNA associations, since the number of individuals with the disease was low among the YFS participants. It might be useful to repeat the current study in a cohort with a higher prevalence of T2DM, and especially in a population harbouring haplogroup N9a, which has the strongest evidence of susceptibility to T2DM (Fang et al. 2018).

7 CONCLUSIONS

The present study demonstrated that mtDNA variants and haplogroups associate with peripheral blood transcriptomics and DNA methylation. Some of the associated transcripts and CpG sites may be linked back to the biological processes taking place in mitochondria, which suggests that the associations may represent the retrograde response in order to maintain cellular homeostasis. However, the majority of the identified CpG sites did not associate with mRNA transcript levels, and the functional relevance of these associations remains unclear. Future studies should verify the causal role of mtDNA variation and also explore the molecular mechanisms behind these reported associations.

We did not observe similar variant associations with transcriptomics and methylation. This indicates that, if there is a causal relationship between mtDNA variation and peripheral blood transcriptomics, the regulatory mechanisms are not mediated by changes in DNAm. Mitochondria affect DNAm mainly via TCA cycle metabolites, whereas nuclear gene expression may also be regulated by mitochondria-derived peptides, Ca^{2+} released from the mitochondrial matrix, or other mitochondrial metabolites, such as ATP and ROS.

This study did not provide proper reassurance for the hypothesis that maternal mitochondrial inheritance leads to sex-specific mitochondrial genetic effects. However, the peripheral blood traits tested for the 'mother's curse' in our natural populations may not have been the most appropriate. We found one mtSNP–transcript pair showing prediabetes-specific dimorphism with plausible functional relevance based on previous studies. A note of caution, however, is that we did not have a replication sample and the possibility of a false positive result should be acknowledged.

Our results, in agreement with several previous studies, suggest that mtDNA variation does not have a significant role in the regulation of blood pressure. In addition, we did not observe any sex-specific effects.

Taken together, this thesis provides both novel and additional evidence in support

of the mitochondrial genetic control of peripheral blood transcriptomics and DNA methylation, with little indication found for sex- and prediabetes-specific effects and no sign of an association between blood pressure and mtDNA variation.

REFERENCES

Achilli, A. et al. (2011). Mitochondrial DNA Backgrounds Might Modulate Diabetes Complications Rather than T2DM as a Whole. *PLOS ONE* 6.6, e21029. DOI: 10.1371/journal.pone.0021029.

Akbari, M., Keijzers, G., Maynard, S., Scheibye-Knudsen, M., Desler, C., Hickson, I. D. and Bohr, V. A. (2014). Overexpression of DNA Ligase III in Mitochondria Protects Cells against Oxidative Stress and Improves Mitochondrial DNA Base Excision Repair. *DNA Repair* 16, 44–53. DOI: 10.1016/j.dnarep.2014.01.015.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215.3, 403–410. DOI: 10.1016/S0022-2836(05)80360-2.

American Diabetes Association (2014). Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 37, S81–S90. DOI: 10.2337/dc14-S081.

Anastasiadi, D., Esteve-Codina, A. and Piferrer, F. (2018). Consistent Inverse Correlation between DNA Methylation of the First Intron and Gene Expression across Tissues and Species. *Epigenetics Chromatin* 11, 37. DOI: 10.1186/s13072-018-0205-1.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. and Young, I. G. (1981). Sequence and Organization of the Human Mitochondrial Genome. *Nature* 290.5806, 457–465. DOI: 10.1038/290457a0.

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. *Nat. Genet.* 23.2, 147–147. DOI: 10.1038/13779.

Arning, L., Haghikia, A., Taherzadeh-Fard, E., Saft, C., Andrich, J., Pula, B., Höxtermann, S., Wieczorek, S., Akkad, D. A., Perrech, M., Gold, R., Epplen, J. T. and Chan, A. (2010). Mitochondrial Haplogroup H Correlates with ATP Levels and Age at Onset in Huntington Disease. *J. Mol. Med.* 88.4, 431–436. DOI: 10.1007/s00109-010-0589-2.

Ashar, F. N., Zhang, Y., Longchamps, R. J., Lane, J., Moes, A., Grove, M. L., Mychaleckyj, J. C., Taylor, K. D., Coresh, J., Rotter, J. I., Boerwinkle, E., Pankratz, N., Guallar, E. and Arking, D. E. (2017). Association of Mitochondrial DNA Copy Number With Cardiovascular Disease. *JAMA Cardiol.* 2.11, 1247–1255. DOI: 10.1001/jamacardio.2017.3683.

Atilano, S. R., Malik, D., Chwa, M., Cáceres-del-Carpio, J., Nesburn, A. B., Boyer, D. S., Kuppermann, B. D., Jazwinski, S. M., Miceli, M. V., Wallace, D. C., Udar, N. and Kenney, M. C. (2015). Mitochondrial DNA Variants Can Mediate Methylation Status of Inflammation, Angiogenesis and Signaling Genes. *Hum. Mol. Genet.* 24.16, 4491–4503. DOI: 10.1093/hmg/ddv173.

Auer, P. L., Reiner, A. P. and Leal, S. M. (2016). The Effect of Phenotypic Outliers and Non-Normality on Rare-Variant Association Testing. *Eur. J. Hum. Genet.* 24.8, 1188–1194. DOI: 10.1038/ejhg.2015.270.

Baccarelli, A. A. and Byun, H.-M. (2015). Platelet Mitochondrial DNA Methylation: A Potential New Marker of Cardiovascular Disease. *Clin. Epigenetics* 7, 44. DOI: 10.1186/s13148-015-0078-0.

Bai, R.-K., Leal, S. M., Covarrubias, D., Liu, A. and Wong, L.-J. C. (2007). Mitochondrial Genetic Background Modifies Breast Cancer Risk. *Cancer Res.* 67.10, 4687–4694. DOI: 10.1158/0008-5472.CAN-06-3554.

Bandelt, H.-J., Kloss-Brandstätter, A., Richards, M. B., Yao, Y.-G. and Logan, I. (2014). The Case for the Continuing Use of the Revised Cambridge Reference Sequence (rCRS) and the Standardization of Notation in Human Mitochondrial DNA Studies. *J. Hum. Genet.* 59.2, 66–77. DOI: 10.1038/jhg.2013.120.

Barrett, J. C., Fry, B., Maller, J. and Daly, M. J. (2005). Haploview: Analysis and Visualization of LD and Haplotype Maps. *Bioinformatics* 21.2, 263–265. DOI: 10.1093/bioinformatics/bth457.

Baudouin, S. V., Saunders, D., Tiangyou, W., Elson, J. L., Poynter, J., Pyle, A., Keers, S., Turnbull, D. M., Howell, N. and Chinnery, P. F. (2005). Mitochondrial DNA

and Survival after Sepsis: A Prospective Study. *Lancet* 366.9503, 2118–2121. DOI: 10.1016/S0140-6736(05)67890-7.

Beekman, M., Dowling, D. K. and Aanen, D. K. (2014). The Costs of Being Male: Are There Sex-Specific Effects of Uniparental Mitochondrial Inheritance?: *Philos. Trans. R. Soc. B* 369.1646, 20130440. DOI: 10.1098/rstb.2013.0440.

Behar, D. M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N. M., Kivisild, T., Torroni, A. and Villems, R. (2012). A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from Its Root. *Am. J. Hum. Genet.* 90.4, 675–684. DOI: 10.1016/j.ajhg.2012.03.002.

Bellizzi, D., D'Aquila, P., Giordano, M., Montesanto, A. and Passarino, G. (2012). Global DNA Methylation Levels Are Modulated by Mitochondrial DNA Variants. *Epigenomics* 4.1, 17–27. DOI: 10.2217/epi.11.109.

Benowitz, N. L., Bernert, J. T., Foulds, J., Hecht, S. S., Jacob III, P., Jarvis, M. J., Joseph, A., Oncken, C. and Piper, M. E. (2020). Biochemical Verification of Tobacco Use and Abstinence: 2019 Update. *Nicotine Tob. Res.* 22.7, 1086–1097. DOI: 10.1093/ntr/ntz132.

Biffi, A., Anderson, C. D., Nalls, M. A., Rahman, R., Sonni, A., Cortellini, L., Rost, N. S., Matarin, M., Hernandez, D. G., Plourde, A., De Bakker, P. I. W., Ross, O. A., Greenberg, S. M., Furie, K. L., Meschia, J. F., Singleton, A. B., Saxena, R. and Rosand, J. (2010). Principal-Component Analysis for Assessment of Population Stratification in Mitochondrial Medical Genetics. *Am. J. Hum. Genet.* 86.6, 904–917. DOI: 10.1016/j.ajhg.2010.05.005.

Birky, C. W. (1994). Relaxed and Stringent Genomes: Why Cytoplasmic Genes Don't Obey Mendel's Laws. *J. Hered.* 85.5, 355–365. DOI: 10.1093/oxfordjournals.jhered.a111480.

Brandes, R. P. (2014). Endothelial Dysfunction and Hypertension. *Hypertension* 64.5, 924–928. DOI: 10.1161/HYPERTENSIONAHA.114.03575.

Brown, W. M., George, M. and Wilson, A. C. (1979). Rapid Evolution of Animal Mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* 76.4, 1967–1971. DOI: 10.1073/pnas.76.4.1967.

Buford, T. W., Todd, M. M., Kairalla, J. A., Mcdermott, M. M., Fragoso, C. A. V. F., Chen, H., Fielding, R. A., King, A. C., Newman, A. B. and Tranah, G. J. (2018). Mitochondrial DNA Sequence Variants Associated with Blood Pressure Among

2 Cohorts of Older Adults. *J. Am. Heart Assoc.* 7.18, e010009. DOI: 10.1161/JAHA.118.010009.

Campbell, S. L. and Wellen, K. E. (2018). Metabolic Signaling to the Nucleus in Cancer. *Mol. Cell* 71.3, 398–408. DOI: 10.1016/j.molcel.2018.07.015.

Cann, R. L., Stoneking, M. and Wilson, A. C. (1987). Mitochondrial DNA and Human Evolution. *Nature* 325.6099, 31–36. DOI: 10.1038/325031a0.

Carey, B. W., Finley, L. W. S., Cross, J. R., Allis, C. D. and Thompson, C. B. (2015). Intracellular Alpha-Ketoglutarate Maintains the Pluripotency of Embryonic Stem Cells. *Nature* 518.7539, 413–416. DOI: 10.1038/nature13981.

Chen, H., Zheng, J., Xue, L., Meng, Y., Wang, Y., Zheng, B., Fang, F., Shi, S., Qiu, Q., Jiang, P., Lu, Z., Mo, J. Q., Lu, J. and Guan, M.-X. (2012). The 12S rRNA A1555G Mutation in the Mitochondrial Haplogroup D5a Is Responsible for Maternally Inherited Hypertension and Hearing Loss in Two Chinese Pedigrees. *Eur. J. Hum. Genet.* 20.6, 607–612. DOI: 10.1038/ejhg.2011.259.

Chen, X., Zhang, Y., Xu, B., Cai, Z., Wang, L., Tian, J., Liu, Y. and Li, Y. (2016). The Mitochondrial Calcium Uniporter Is Involved in Mitochondrial Calcium Cycle Dysfunction: Underlying Mechanism of Hypertension Associated with Mitochondrial tRNAIle A4263G Mutation. *Int. J. Biochem. Cell B.* 78, 307–314. DOI: 10.1016/j.biocel.2016.07.018.

Chinnery, P. F., Elliott, H. R., Patel, S., Lambert, C., Keers, S. M., Durham, S. E., McCarthy, M. I., Hitman, G. A., Hattersley, A. T. and Walker, M. (2005). Role of the Mitochondrial DNA 16184-16193 Poly-C Tract in Type 2 Diabetes. *Lancet* 366.9497, 1650–1651. DOI: 10.1016/S0140-6736(05)67492-2.

Chinnery, P. F. and Hudson, G. (2013). Mitochondrial Genetics. *Br. Med. Bull.* 106.1, 135–159. DOI: 10.1093/bmb/ldt017.

Chinnery, P. F., Mowbray, C., Patel, S. K., Elson, J. L., Sampson, M., Hitman, G. A., McCarthy, M. I., Hattersley, A. T. and Walker, M. (2007). Mitochondrial DNA Haplogroups and Type 2 Diabetes: A Study of 897 Cases and 1010 Controls. *J. Med. Genet.* 44.6, e80. DOI: 10.1136/jmg.2007.048876.

Chinnery, P. F. and Schon, E. A. (2003). Mitochondria. *J. Neurol. Neurosurg. Psychiatry* 74.9, 1188–1199. DOI: 10.1136/jnnp.74.9.1188.

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. and Turnbull, D. M. (2004). Risk of Developing a Mitochondrial DNA Deletion Disorder. *Lancet* 364.9434, 592–596. DOI: 10.1016/S0140-6736(04)16851-7.

Chinnery, P. F., Elliott, H. R., Syed, A. and Rothwell, P. M. (2010). Mitochondrial DNA Haplogroups and Risk of Transient Ischaemic Attack and Ischaemic Stroke: A Genetic Association Study. *Lancet Neurol.* 9.5, 498–503. DOI: 10.1016/S1474-4422(10)70083-1.

Chinnery, P. F., Elliott, H. R., Hudson, G., Samuels, D. C. and Relton, C. L. (2012). Epigenetics, Epidemiology and Mitochondrial DNA Diseases. *Int. J. Epidemiol.* 41.1, 177–187. DOI: 10.1093/ije/dyr232.

Cohen, T., Levin, L. and Mishmar, D. (2016). Ancient Out-of-Africa Mitochondrial DNA Variants Associate with Distinct Mitochondrial Gene Expression Patterns. *PLOS Genet.* 12.11, e1006407. DOI: 10.1371/journal.pgen.1006407.

Cortés-Pereira, E., Fernández-Tajes, J., Fernández-Moreno, M., Vázquez-Mosquera, M. E., Relaño, S., Ramos-Louro, P., Durán-Sotuela, A., Dalmao-Fernández, A., Oreiro, N., Blanco, F. J. and Rego-Pérez, I. (2019). Differential Association of Mitochondrial DNA Haplogroups J and H With the Methylation Status of Articular Cartilage: Potential Role in Apoptosis and Metabolic and Developmental Processes. *Arthritis Rheumatol.* 71.7, 1191–1200. DOI: 10.1002/art.40857.

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. (2008). A Reduction of Mitochondrial DNA Molecules during Embryogenesis Explains the Rapid Segregation of Genotypes. *Nat. Genet.* 40.2, 249–254. DOI: 10.1038/ng.2007.63.

Cui, J. S., Hopper, J. L. and Harrap, S. B. (2003). Antihypertensive Treatments Obscure Familial Contributions to Blood Pressure Variation. *Hypertension* 41.2, 207–210. DOI: 10.1161/01.HYP.0000044938.94050.E3.

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.* 42.20, 12640–12649. DOI: 10.1093/nar/gku1038.

De Bakker, P. I. W., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J. and Altshuler, D. (2005). Efficiency and Power in Genetic Association Studies. *Nat. Genet.* 37.11, 1217–1223. DOI: 10.1038/ng1669.

Ding, Y., Xia, B., Yu, J., Leng, J. and Huang, J. (2013). Mitochondrial DNA Mutations and Essential Hypertension (Review). *Int. J. Mol. Med.* 32.4, 768–774. DOI: 10.3892/ijmm.2013.1459.

Dor, Y. and Cedar, H. (2018). Principles of DNA Methylation and Their Implications for Biology and Medicine. *Lancet* 392.10149, 777–786. DOI: 10.1016/S0140-6736(18)31268-6.

Dowling, D. K. and Adrian, R. E. (2019). Challenges and Prospects for Testing the Mother's Curse Hypothesis. *Integr. Comp. Biol.* 59.4, 875–889. DOI: 10.1093/icb/icz110.

Dunham-Snary, K. J., Sandel, M. W., Sammy, M. J., Westbrook, D. G., Xiao, R., McMonigle, R. J., Ratcliffe, W. F., Penn, A., Young, M. E. and Ballinger, S. W. (2018). Mitochondrial – Nuclear Genetic Interaction Modulates Whole Body Metabolism, Adiposity and Gene Expression in Vivo. *EBioMedicine* 36, 316–328. DOI: 10.1016/j.ebiom.2018.08.036.

Ebner, S., Mangge, H., Langhof, H., Halle, M., Siegrist, M., Aigner, E., Paulmichl, K., Paulweber, B., Datz, C., Sperl, W., Kofler, B. and Weghuber, D. (2015). Mitochondrial Haplogroup T Is Associated with Obesity in Austrian Juveniles and Adults. *PLOS ONE* 10.8, e0135622. DOI: 10.1371/journal.pone.0135622.

Ehret, G. B. et al. (2011). Genetic Variants in Novel Pathways Influence Blood Pressure and Cardiovascular Disease Risk. *Nature* 478.7367, 103–109. DOI: 10.1038/nature10405.

English, J., Son, J. M., Cardamone, M. D., Lee, C. and Perissi, V. (2020). Decoding the Rosetta Stone of Mitonuclear Communication. *Pharmacol. Res.* 161, 105161. DOI: 10.1016/j.phrs.2020.105161.

Fang, H., Hu, N., Zhao, Q., Wang, B., Zhou, H., Fu, Q., Shen, L., Chen, X., Shen, F. and Lyu, J. (2018). mtDNA Haplogroup N9a Increases the Risk of Type 2 Diabetes by Altering Mitochondrial Function and Intracellular Mitochondrial Signals. *Diabetes* 67.7, 1441–1453. DOI: 10.2337/db17-0974.

Fernández-Moreno, M., Soto-Hermida, A., Vázquez-Mosquera, M. E., Cortés-Pereira, E., Relaño, S., Hermida-Gómez, T., Pértega, S., Oreiro-Villar, N., Fernández-López, C., Garesse, R., Blanco, F. J. and Rego-Pérez, I. (2017). Mitochondrial DNA Haplogroups Influence the Risk of Incident Knee Osteoarthritis in OAI and CHECK Cohorts. A Meta-Analysis and Functional Study. *Ann. Rheum. Dis.* 76.6, 1114–1122. DOI: 10.1136/annrheumdis-2016-210131.

Fetterman, J. L. and Ballinger, S. W. (2019). Mitochondrial Genetics Regulate Nuclear Gene Expression through Metabolites. *Proc. Natl. Acad. Sci.* 116.32, 15763–15765. DOI: 10.1073/pnas.1909996116.

Fetterman, J. L., Zelickson, B. R., Johnson, L. W., Moellering, D. R., Westbrook, D. G., Pompilius, M., Sammy, M. J., Johnson, M., Dunham-Snary, K. J., Cao, X., Bradley, W. E., Zhang, J., Wei, C.-C., Chacko, B., Schurr, T. G., Kesterson, R. A., Dell'italia, L. J., Darley-Usmar, V. M., Welch, D. R. and Ballinger, S. W. (2013). Mitochondrial Genetic Background Modulates Bioenergetics and Susceptibility to Acute Cardiac Volume Overload. *Biochem. J.* 455.2, 157–167. DOI: 10.1042/BJ20130029.

Fiorito, G. et al. (2019). Socioeconomic Position, Lifestyle Habits and Biomarkers of Epigenetic Aging: A Multi-Cohort Analysis. *Aging* 11.7, 2045–2070. DOI: 10.18632/aging.101900.

Fish, J., Raule, N. and Attardi, G. (2004). Discovery of a Major D-Loop Replication Origin Reveals Two Modes of Human mtDNA Synthesis. *Science* 306.5704, 2098–2101. DOI: 10.1126/science.1102077.

Formentini, L., Sánchez-Aragó, M., Sánchez-Cenizo, L. and Cuezva, J. M. (2012). The Mitochondrial ATPase Inhibitory Factor 1 Triggers a ROS-Mediated Retrograde Prosurvival and Proliferative Response. *Mol. Cell* 45.6, 731–742. DOI: 10.1016/j.molcel.2012.01.008.

Fortin, J.-P., Triche Jr, T. J. and Hansen, K. D. (2017). Preprocessing, Normalization and Integration of the Illumina HumanMethylationEPIC Array with Minfi. *Bioinformatics* 33.4, 558–560. DOI: 10.1093/bioinformatics/btw691.

Frank, S. A. and Hurst, S. A. (1996). Mitochondria and Male Disease. *Nature* 383.6597, 224–224. DOI: 10.1038/383224a0.

Friedman, J. R. and Nunnari, J. (2014). Mitochondrial Form and Function. *Nature* 505.7483, 335–343. DOI: 10.1038/nature12985.

Friedrich, V. K., Rubel, M. A. and Schurr, T. G. (2021). Mitochondrial Genetic Variation in Human Bioenergetics, Adaptation, and Adult Disease. *Am. J. Hum. Biol.* 34.2, e23629. DOI: 10.1002/ajhb.23629.

Fuku, N., Park, K. S., Yamada, Y., Nishigaki, Y., Cho, Y. M., Matsuo, H., Segawa, T., Watanabe, S., Kato, K., Yokoi, K., Nozawa, Y., Lee, H. K. and Tanaka, M. (2007). Mitochondrial Haplogroup N9a Confers Resistance against Type 2 Diabetes in Asians. *Am. J. Hum. Genet.* 80.3, 407–415. DOI: 10.1086/512202.

Garcia-Roves, P. M., Osler, M. E., Holmström, M. H. and Zierath, J. R. (2008). Gain-of-Function R225Q Mutation in AMP-activated Protein Kinase Gamma-3 Subunit Increases Mitochondrial Biogenesis in Glycolytic Skeletal Muscle. *J. Biol. Chem.* 283.51, 35724–35734. DOI: 10.1074/jbc.M805078200.

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 Ecol. Evol.* 19.5, 238–244. DOI: 10.1016/j.tree.2004.02.002.

Ghosh, A. R., Bhattacharya, R., Bhattacharya, S., Nargis, T., Rahaman, O., Duttagupta, P., Raychaudhuri, D., Shiu, C., Liu, C., Roy, S., Ghosh, P., Khanna, S., Chaudhuri, T., Tantia, O., Haak, S., Bandyopadhyay, S., Mukhopadhyay, S., Chakrabarti, P. and Ganguly, D. (2016). Adipose Recruitment and Activation of Plasmacytoid Dendritic Cells Fuel Metaflammation. *Diabetes* 65.11, 3440–3452. DOI: 10.2337/db16-0331.

Giacomello, M., Pyakurel, A., Glytsou, C. and Scorrano, L. (2020). The Cell Biology of Mitochondrial Membrane Dynamics. *Nat. Rev. Mol. Cell Biol.* 21.4, 204–224. DOI: 10.1038/s41580-020-0210-7.

Giles, R. E., Blanc, H., Cann, H. M. and Wallace, D. C. (1980). Maternal Inheritance of Human Mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* 77.11, 6715–6719. DOI: 10.1073/pnas.77.11.6715.

Gomez-Alonso, M. d. C. et al. (2021). DNA Methylation and Lipid Metabolism: An EWAS of 226 Metabolic Measures. *Clin. Epigenetics* 13, 7. DOI: 10.1186/s13148-020-00957-8.

González, M. d. M., Ramos, A., Aluja, M. P. and Santos, C. (2020). Sensitivity of Mitochondrial DNA Heteroplasmy Detection Using Next Generation Sequencing. *Mitochondrion* 50, 88–93. DOI: 10.1016/j.mito.2019.10.006.

Gorman, G. S., Schaefer, A. M., Ng, Y., Gomez, N., Blakely, E. L., Alston, C. L., Feeney, C., Horvath, R., Yu-Wai-Man, P., Chinnery, P. F., Taylor, R. W., Turnbull, D. M. and McFarland, R. (2015). Prevalence of Nuclear and Mitochondrial DNA Mutations Related to Adult Mitochondrial Disease. *Ann. Neurol.* 77.5, 753–759. DOI: 10.1002/ana.24362.

Gray, M. W., Burger, G. and Lang, B. F. (1999). Mitochondrial Evolution. *Science* 283.5407, 1476–1481. DOI: 10.1126/science.283.5407.1476.

Greaves, L. C., Reeve, A. K., Taylor, R. W. and Turnbull, D. M. (2012). Mitochondrial DNA and Disease. *J. Pathol.* 226.2, 274–286. DOI: 10.1002/path.3028.

Greenberg, M. V. C. and Bourc'his, D. (2019). The Diverse Roles of DNA Methylation in Mammalian Development and Disease. *Nat. Rev. Mol. Cell Biol.* 20.10, 590–607. DOI: 10.1038/s41580-019-0159-6.

Guha, M. and Avadhani, N. G. (2013). Mitochondrial Retrograde Signaling at the Crossroads of Tumor Bioenergetics, Genetics and Epigenetics. *Mitochondrion* 13.6, 577–591. DOI: 10.1016/j.mito.2013.08.007.

Hauswirth, W. W. and Laipis, P. J. (1982). Mitochondrial DNA Polymorphism in a Maternal Lineage of Holstein Cows. *Proc. Natl. Acad. Sci. U.S.A.* 79.15, 4686–4690. DOI: 10.1073/pnas.79.15.4686.

Hazafa, A., Batool, A., Ahmad, S., Amjad, M., Chaudhry, S. N., Asad, J., Ghuman, H. F., Khan, H. M., Naeem, M. and Ghani, U. (2021). Humanin: A Mitochondrial-Derived Peptide in the Treatment of Apoptosis-Related Diseases. *Life Sci.* 264, 118679. DOI: 10.1016/j.lfs.2020.118679.

Hedman, M., Brandstätter, A., Pimenoff, V., Sistonen, P., Palo, J. U., Parson, W. and Sajantila, A. (2007). Finnish Mitochondrial DNA HVS-I and HVS-II Population Data. *Forensic Sci. Int.* 172.2, 171–178. DOI: 10.1016/j.forsciint.2006.09.012.

Hertzog Santos, J. (2021). Mitochondria Signaling to the Epigenome: A Novel Role for an Old Organelle. *Free Radical Biology and Medicine* 170, 59–69. DOI: 10.1016/j.freeradbiomed.2020.11.016.

Hinttala, R., Kervinen, M., Uusimaa, J., Maliniemi, P., Finnilä, S., Rantala, H., Remes, A. M., Hassinen, I. E. and Majamaa, K. (2010). Analysis of Functional Consequences of Haplogroup J Polymorphisms m.4216T>C and m.3866T>C in Human MT-ND1: Mutagenesis of Homologous Positions in Escherichia Coli. *Mitochondrion* 10.4, 358–361. DOI: 10.1016/j.mito.2010.02.002.

Holyoake, A. J., McHugh, P., Wu, M., O'Carroll, S., Benny, P., Sin, I. L. and Sin, F. Y. T. (2001). High Incidence of Single Nucleotide Substitutions in the Mitochondrial Genome Is Associated with Poor Semen Parameters in Men. *Int. J. Androl.* 24.3, 175–182. DOI: 10.1046/j.1365-2605.2001.00292.x.

Horvath, S. (2013). DNA Methylation Age of Human Tissues and Cell Types. *Genome Biol.* 14.10, 3156. DOI: 10.1186/gb-2013-14-10-r115.

Hostalek, U. (2019). Global Epidemiology of Prediabetes - Present and Future Perspectives. *Clin. Diabetes Endocrinol.* 5.1, 5. DOI: 10.1186/s40842-019-0080-0.

Hottenga, J.-J., Whitfield, J. B., Geus, E. J. C. de, Boomsma, D. I. and Martin, N. G. (2006). Heritability and Stability of Resting Blood Pressure in Australian Twins. *Twin Res. Hum. Genet.* 9.2, 205–209. DOI: 10.1375/twin.9.2.205.

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, 86. DOI: 10.1186/1471-2105-13-86.

Huan, T., Joehanes, R., Song, C., Peng, F., Guo, Y., Mendelson, M., Yao, C., Liu, C., Ma, J., Richard, M., Agha, G., Guan, W., Almli, L. M., Conneely, K. N., Keefe, J., Hwang, S.-J., Johnson, A. D., Fornage, M., Liang, L. and Levy, D. (2019). Genome-Wide Identification of DNA Methylation QTLs in Whole Blood Highlights Pathways for Cardiovascular Disease. *Nat. Commun.* 10.1, 4267. DOI: 10.1038/s41467-019-12228-z.

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. (2013). Two-Stage Association Study and Meta-Analysis of Mitochondrial DNA Variants in Parkinson Disease. *Neurology* 80.22, 2042–2048. DOI: 10.1212/WNL.0b013e318294b434.

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 Genet.* 10.5, e1004369. DOI: 10.1371/journal.pgen.1004369.

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, e0163770. DOI: 10.1371/journal.pone.0163770.

Hutchison, C. A., Newbold, J. E., Potter, S. S. and Edgell, M. H. (1974). Maternal Inheritance of Mammalian Mitochondrial DNA. *Nature* 251.5475, 536–538. DOI: 10.1038/251536a0.

Hwang, S., Kwak, S. H., Bhak, J., Kang, H. S., Lee, Y. R., Koo, B. K., Park, K. S., Lee, H. K., Cho, Y. M., Heon Kwak, S., Bhak, J., Kang, H. S., Lee, Y. R., Koo, B. K., Park, K. S., Lee, H. K. and Cho, M. (2011). Gene Expression Pattern in Transmitochondrial Cytoplasmic Hybrid Cells Harboring Type 2 Diabetes-Associated

Mitochondrial DNA Haplogroups. *PLOS ONE* 6.7, e22116. DOI: 10.1371/journal.pone.0022116.

Ingman, M. and Gyllensten, U. (2007). A Recent Genetic Link between Sami and the Volga-Ural Region of Russia. *Eur. J. Hum. Genet.* 15.1, 115–120. DOI: 10.1038/sj.ejhg.5201712.

Ingman, M., Kaessmann, H., Pääbo, S. and Gyllensten, U. (2000). Mitochondrial Genome Variation and the Origin of Modern Humans. *Nature* 408.6813, 708–713. DOI: 10.1038/35047064.

Innocenti, P., Morrow, E. H. and Dowling, D. K. (2011). Experimental Evidence Supports a Sex-Specific Selective Sieve in Mitochondrial Genome Evolution. *Science* 332.6031, 845–848.

Ioannidis, J. P. A. (2007). Non-Replication and Inconsistency in the Genome-Wide Association Setting. *Hum. Hered.* 64.4, 203–213. DOI: 10.1159/000103512.

Jansen, R., Batista, S., Brooks, A. I., Tischfield, J. A., Willemsen, G., van Grootheest, G., Hottenga, J.-J., Milaneschi, Y., Mbarek, H., Madar, V., Peyrot, W., Vink, J. M., Verweij, C. L., de Geus, E. J., Smit, J. H., Wright, F. A., Sullivan, P. F., Boomsma, D. I. and Penninx, B. W. (2014). Sex Differences in the Human Peripheral Blood Transcriptome. *BMC Genomics* 15, 33. DOI: 10.1186/1471-2164-15-33.

Juvinao-Quintero, D. L., Marioni, R. E., Ochoa-Rosales, C., Russ, T. C., Deary, I. J., van Meurs, J. B. J., Voortman, T., Hivert, M.-F., Sharp, G. C., Relton, C. L. and Elliott, H. R. (2021). DNA Methylation of Blood Cells Is Associated with Prevalent Type 2 Diabetes in a Meta-Analysis of Four European Cohorts. *Clin. Epigenetics* 13, 40. DOI: 10.1186/s13148-021-01027-3.

Kassam, I., Qi, T., Lloyd-Jones, L., Holloway, A., Bonder, M. J., Henders, A. K., Martin, N. G., Powell, J. E., Franke, L., Montgomery, G. W., Visscher, P. M. and McRae, A. F. (2016). Evidence for Mitochondrial Genetic Control of Autosomal Gene Expression. *Hum. Mol. Genet.* 25.24, 5332–5338. DOI: 10.1093/hmg/ddw347.

Kenney, M. C., 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. (2014a). Inherited Mitochondrial DNA Variants Can Affect Complement, In-

flammation and Apoptosis Pathways: Insights into Mitochondrial-Nuclear Interactions. *Hum. Mol. Genet.* 23.13, 3537–3551. DOI: 10.1093/hmg/ddu065.

Kenney, M. C., Chwa, M., Atilano, S. R., Falatoonzadeh, P., Ramirez, C., Malik, D., Tarek, M., del Carpio, J. C., Nesburn, A. B., Boyer, D. S., Kuppermann, B. D., Vawter, M. P., Jazwinski, S. M., Miceli, M. V., Wallace, D. C. and Udar, N. (2014b). Molecular and Bioenergetic Differences between Cells with African versus European Inherited Mitochondrial DNA Haplogroups: Implications for Population Susceptibility to Diseases. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1842.2, 208–219. DOI: 10.1016/j.bbadis.2013.10.016.

Kenney, M. C., Chwa, M., Atilano, S. R., Pavlis, J. M., Falatoonzadeh, P., Ramirez, C., Malik, D., Hsu, T., Woo, G., Soe, K., Nesburn, A. B., Boyer, D. S., Kuppermann, B. D., Jazwinski, S. M., Miceli, M. V., Wallace, D. C. and Udar, N. (2013). Mitochondrial DNA Variants Mediate Energy Production and Expression Levels for CFH, C3 and EFEMP1 Genes: Implications for Age-Related Macular Degeneration. *PLOS ONE* 8.1, e54339. DOI: 10.1371/journal.pone.0054339.

Khramtsova, E. A., Davis, L. K. and Stranger, B. E. (2019). The Role of Sex in the Genomics of Human Complex Traits. *Nat. Rev. Genet.* 20.3, 173–190. DOI: 10.1038/s41576-018-0083-1.

Kim, S.-J., Xiao, J., Wan, J., Cohen, P. and Yen, K. (2017). Mitochondrially Derived Peptides as Novel Regulators of Metabolism. *J. Physiol.* 595.21, 6613–6621. DOI: 10.1113/JP274472.

Kim, K. H., Son, J. M., Benayoun, B. A. and Lee, C. (2018). The Mitochondrial-Encoded Peptide MOTS-c Translocates to the Nucleus to Regulate Nuclear Gene Expression in Response to Metabolic Stress. *Cell Metab.* 28.3, 516–524.e7. DOI: 10.1016/j.cmet.2018.06.008.

Kim-Han, J. S., Reichert, S. A., Quick, K. L. and Dugan, L. L. (2001). BMCP1: A Mitochondrial Uncoupling Protein in Neurons Which Regulates Mitochondrial Function and Oxidant Production. *J. Neurochem.* 79.3, 658–668. DOI: 10.1046/j.1471-4159.2001.00604.x.

Kirkman, M. A., Yu-Wai-Man, P., Korsten, A., Leonhardt, M., Dimitriadis, K., De Coo, I. F., Klopstock, T. and Chinnery, P. F. (2009). Gene–Environment Interactions in Leber Hereditary Optic Neuropathy. *Brain* 132.9, 2317–2326. DOI: 10.1093/brain/awp158.

Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K. K., Passarino, G., Underhill, P. A., Scharfe, C., Torroni, A., Scozzari, R., Modiano, D., Coppa, A., de Knijff, P., Feldman, M., Cavalli-Sforza, L. L. and Oefner, P. J. (2005). The Role of Selection in the Evolution of Human Mitochondrial Genomes. *Genetics* 172.1, 373–387. DOI: 10.1534/genetics.105.043901.

Kopinski, P. K., Janssen, K. A., Schaefer, P. M., Trefely, S., Perry, C. E., Potluri, P., Tintos-Hernandez, J. A., Singh, L. N., Karch, K. R., Campbell, S. L., Doan, M. T., Jiang, H., Nissim, I., Nakamaru-Ogiso, E., Wellen, K. E., Snyder, N. W., Garcia, B. A. and Wallace, D. C. (2019). Regulation of Nuclear Epigenome by Mitochondrial DNA Heteroplasmy. *Proc. Natl. Acad. Sci. U.S.A.* 116.32, 16028–16035. DOI: 10.1073/pnas.1906896116.

Kraja, A. T. et al. (2019). Associations of Mitochondrial and Nuclear Mitochondrial Variants and Genes with Seven Metabolic Traits. *Am. J. Hum. Genet.* 104.1, 112–138. DOI: 10.1016/J.AJHG.2018.12.001.

Kriebel, J., Herder, C., Rathmann, W., Wahl, S., Kunze, S. and Molnos, S. (2016). Association between DNA Methylation in Whole Blood and Measures of Glucose Metabolism: KORA F4 Study. *PLOS ONE* 11.3, e0152314. DOI: 10.1371/journal.pone.0152314.

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?: *Nat. Genet.* 40.3, 275–279. DOI: 10.1038/ng.f.94.

Kwok, K. H.-H., Ho, P. W.-L., Chu, A. C.-Y., Ho, J. W.-M., Liu, H.-F., Yiu, D. C.-W., Chan, K.-H., Kung, M. H.-W., Ramsden, D. B. and Ho, S.-L. (2010). Mitochondrial UCP5 Is Neuroprotective by Preserving Mitochondrial Membrane Potential, ATP Levels, and Reducing Oxidative Stress in MPP+ and Dopamine Toxicity. *Free Radical Bio. Med.* 49.6, 1023–1035. DOI: 10.1016/j.freeradbiomed.2010.06.017.

Lakshmipathy, U. and Campbell, C. (1999). The Human DNA Ligase III Gene Encodes Nuclear and Mitochondrial Proteins. *Mol. Cell. Biol.* 19.5, 3869–3876. DOI: 10.1128/MCB.19.5.3869.

Landgraf, A. J. and Lee, Y. (2020). Dimensionality Reduction for Binary Data through the Projection of Natural Parameters. *J. Multivar. Anal.* 180, 104668. DOI: 10.1016/j.jmva.2020.104668.

Laricchia, K. M., Lake, N. J., Watts, N. A., Shand, M., Haessly, A., Gauthier, L., Benjamin, D., Banks, E., Soto, J., Garimella, K., Emery, J., Genome Aggregation Database Consortium, Rehm, H. L., MacArthur, D. G., Tiao, G., Lek, M. V., Mootha, V. K. and Calvo, S. E. (2022). Mitochondrial DNA Variation across 56,434 Individuals in gnomAD. *Genome Res.* 32.3, 569–582. DOI: 10.1101/gr.276013.121.

Lee, W. T., Sun, X., Tsai, T.-S., Johnson, J. L., Gould, J. A., Garama, D. J., Gough, D. J., McKenzie, M., Trounce, I. A. and John, J. C. S. (2017). Mitochondrial DNA Haplotypes Induce Differential Patterns of DNA Methylation That Result in Differential Chromosomal Gene Expression Patterns. *Cell Death Dis.* 3.1, 1–11. DOI: 10.1038/cddiscovery.2017.62.

Lehne, B., Drong, A. W., Loh, M., Zhang, W., Scott, W. R., Tan, S.-T., Afzal, U., Scott, J., Jarvelin, M.-R., Elliott, P., McCarthy, M. I., Kooner, J. S. and Chambers, J. C. (2015). A Coherent Approach for Analysis of the Illumina HumanMethylation450 BeadChip Improves Data Quality and Performance in Epigenome-Wide Association Studies. *Genome Biol.* 16, 37. DOI: 10.1186/s13059-015-0600-x.

Lemire, M., Zaidi, S. H. E., Ban, M., Ge, B., Aïssi, D., Germain, M., Kassam, I., Wang, M., Zanke, B. W., Gagnon, F., Morange, P.-E., Trégouët, D.-A., Wells, P. S., Sawcer, S., Gallinger, S., Pastinen, T. and Hudson, T. J. (2015). Long-Range Epigenetic Regulation Is Conferred by Genetic Variation Located at Thousands of Independent Loci. *Nat. Commun.* 6.1, 1–12. DOI: 10.1038/ncomms7326.

Li, B. and Leal, S. M. (2008). Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. *Am. J. Hum. Genet.* 83.3, 311–321. DOI: 10.1016/j.ajhg.2008.06.024.

Li, C. I., Mirus, J. E., Zhang, Y., Ramirez, A. B., Ladd, J. J., Prentice, R. L., McIntosh, M. W., Hanash, S. M. and Lampe, P. D. (2012). Discovery and Preliminary Confirmation of Novel Early Detection Biomarkers for Triple-Negative Breast Cancer Using Preclinical Plasma Samples from the Women's Health Initiative Observational Study. *Breast Cancer Res. Treat.* 135.2, 611–618. DOI: 10.1007/s10549-012-2204-4.

Li, H. and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25.14, 1754–1760. DOI: 10.1093/bioinformatics/btp324.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25.16, 2078–2079. DOI: 10.1093/bioinformatics/btp352.

Li, H., Zhang, X., Wang, F., Ling, Z., Zhongwei, Y., Jiahui, F., Xiang, N., Peihua, W., Xiang-Dong, F., Chen, C. and Wen, W. D. (2016). MicroRNA-21 Lowers Blood Pressure in Spontaneous Hypertensive Rats by Upregulating Mitochondrial Translation. *Circulation* 134.10, 734–751. DOI: 10.1161/CIRCULATIONAHA.116.023926.

Li, J. and Ji, L. (2005). Adjusting Multiple Testing in Multilocus Analyses Using the Eigenvalues of a Correlation Matrix. *Heredity* 95.3, 221–227. DOI: 10.1038/sj.hdy.6800717.

Lim, S. S. et al. (2012). A Comparative Risk Assessment of Burden of Disease and Injury Attributable to 67 Risk Factors and Risk Factor Clusters in 21 Regions, 1990-2010: A Systematic Analysis for the Global Burden of Disease Study 2010. *The Lancet* 380.9859, 2224–2260. DOI: 10.1016/S0140-6736(12)61766-8.

Liu, C., Yang, Q., Hwang, S.-J., Sun, F., Johnson, A. D., Shirihai, O. S., Vasan, R. S., Levy, D. and Schwartz, F. (2012). Association of Genetic Variation in the Mitochondrial Genome with Blood Pressure and Metabolic Traits. *Hypertension* 60.4, 949–956. DOI: 10.1161/HYPERTENSIONAHA.112.196519.

Liu, C., Fetterman, J. L., Sun, X., Yan, K., Liu, P., Luo, Y., Ding, J., Zhu, J. and Levy, D. (2022). Comparison of Mitochondrial DNA Sequences from Whole Blood and Lymphoblastoid Cell Lines. *Sci. Rep.* 12.1, 1801. DOI: 10.1038/s41598-022-05814-7.

Liu, C. et al. (2016a). Meta-Analysis Identifies Common and Rare Variants Influencing Blood Pressure and Overlapping with Metabolic Trait Loci. *Nat. Genet.* 48.10, 1162–1170. DOI: 10.1038/ng.3660.

Liu, X. et al. (2021). Association of Mitochondrial DNA Copy Number with Cardiometabolic Diseases. *Cell Genomics* 1.1, 100006. DOI: 10.1016/j.xgen.2021.100006.

Liu, Y., Beyer, A. and Aebersold, R. (2016b). On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* 165.3, 535–550. DOI: 10.1016/j.cell.2016.03.014.

Lopes, A. F. C. (2020). Mitochondrial Metabolism and DNA Methylation: A Review of the Interaction between Two Genomes. *Clin. Epigenetics* 12.1, 182. DOI: 10.1186/s13148-020-00976-5.

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* 44.1, 23–26. DOI: 10.1002/0471250953.bi0123s44.

Lowe, R., Slodkowicz, G., Goldman, N. and Rakyan, V. K. (2015). The Human Blood DNA Methylome Displays a Highly Distinctive Profile Compared with Other Somatic Tissues. *Epigenetics* 10.4, 274–281. DOI: 10.1080/15592294.2014.1003744.

Luo, S., Valencia, C. A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., Wang, X., Li, Z., Dell, S., Brown, J. et al. (2018). Biparental Inheritance of Mitochondrial DNA in Humans. *Proc. Natl. Acad. Sci. U.S.A.* 115.51, 13039–13044. DOI: 10.1073/pnas.1810946115.

Luo, Y., Bond, J. D. and Ingram, V. M. (1997). Compromised Mitochondrial Function Leads to Increased Cytosolic Calcium and to Activation of MAP Kinases. *Proc. Natl. Acad. Sci. U.S.A.* 94.18, 9705–9710. DOI: 10.1073/pnas.94.18.9705.

Lutz-Bonengel, S. and Parson, W. (2019). No Further Evidence for Paternal Leakage of Mitochondrial DNA in Humans Yet. *Proc. Natl. Acad. Sci. U.S.A.* 116.6, 1821–1822. DOI: 10.1073/pnas.1820533116.

Mägi, R., Lindgren, C. M. and Morris, A. P. (2010). Meta-Analysis of Sex-Specific Genome-Wide Association Studies. *Genet. Epidemiol.* 34.8, 846–853. DOI: 10.1002/gepi.20540.

Mägi, R. and Morris, A. P. (2010). GWAMA: Software for Genome-Wide Association Meta-Analysis. *BMC Bioinformatics* 11, 288. DOI: 10.1186/1471-2105-11-288.

Maier, T., Güell, M. and Serrano, L. (2009). Correlation of mRNA and Protein in Complex Biological Samples. *FEBS Lett.* 583.24, 3966–3973. DOI: 10.1016/j.febslet.2009.10.036.

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. *Am. J. Hum. Genet.* 72.2, 333–339. DOI: 10.1086/346066.

Mansell, G., Gorrie-Stone, T. J., Bao, Y., Kumari, M., Schalkwyk, L. S., Mill, J. and Hannon, E. (2019). Guidance for DNA Methylation Studies: Statistical Insights from the Illumina EPIC Array. *BMC Genomics* 20, 366. DOI: 10.1186/s12864-019-5761-7.

Marcuello, A., Martínez-Redondo, D., Dahmani, Y., Casajús, J. A., Ruiz-Pesini, E., Montoya, J., López-Pérez, M. J. and Díez-Sánchez, C. (2009). Human Mitochondrial Variants Influence on Oxygen Consumption. *Mitochondrion* 9.1, 27–30. DOI: 10.1016/j.mito.2008.10.002.

Margueron, R. and Reinberg, D. (2010). Chromatin Structure and the Inheritance of Epigenetic Information. *Nat. Rev. Genet.* 11.4, 285–296. DOI: 10.1038/nrg2752.

Margulis, L. (1970). Recombination of Non-Chromosomal Genes in Chlamydomonas: Assortment of Mitochondria and Chloroplasts?: *J. Theor. Biol.* 26.2, 337–342. DOI: 10.1016/s0022-5193(70)80023-6.

Martínez-Reyes, I. and Chandel, N. S. (2020). Mitochondrial TCA Cycle Metabolites Control Physiology and Disease. *Nat. Commun.* 11.1, 102. DOI: 10.1038/s41467-019-13668-3.

Matilainen, O., Quirós, P. M. and Auwerx, J. (2017). Mitochondria and Epigenetics – Crosstalk in Homeostasis and Stress. *Trends Cell Biol.* 27.6, 453–463. DOI: 10.1016/j.tcb.2017.02.004.

Matsha, T. E., Pheiffer, C., Humphries, S. E., Gamieldien, J., Erasmus, R. T. and Kengne, A. P. (2016). Genome-Wide DNA Methylation in Mixed Ancestry Individuals with Diabetes and Prediabetes from South Africa. *Int. J. Endocrinol.* 2016, 1–11. DOI: 10.1155/2016/3172093.

Melchinger, H., Jain, K., Tyagi, T. and Hwa, J. (2019). Role of Platelet Mitochondria: Life in a Nucleus-Free Zone. *Front. Cardiovasc. Med.* 6, 153.

Melé, M. et al. (2015). The Human Transcriptome across Tissues and Individuals. *Science* 348.6235, 660–665. DOI: 10.1126/science.aaa0355.

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. *Hum. Genet.* 133.9, 1149–1159. DOI: 10.1007/s00439-014-1458-9.

Messerli, F. H., Williams, B. and Ritz, E. (2007). Essential Hypertension. *Lancet* 370.9587, 591–603. DOI: 10.1016/S0140-6736(07)61299-9.

Michels, K. B., Binder, A. M., Dedeurwaerder, S., Epstein, C. B., Greally, J. M., Gut, I., Houseman, E. A., Izzi, B., Kelsey, K. T., Meissner, A., Milosavljevic, A., Siegmund, K. D., Bock, C. and Irizarry, R. A. (2013). Recommendations for the Design and Analysis of Epigenome-Wide Association Studies. *Nat. Methods* 10.10, 949–955. DOI: 10.1038/nmeth.2632.

Middelberg, R. P., Ferreira, M. A., Henders, A. K., Heath, A. C., Madden, P. A., Montgomery, G. W., Martin, N. G. and Whitfield, J. B. (2011). Genetic Variants in LPL, OASL and TOMM40/APOE-C1-C2-C4 Genes Are Associated with Multiple Cardiovascular-Related Traits. *BMC Med. Genet.* 12, 123. DOI: 10.1186/1471-2350-12-123.

Miller, B., Arpawong, T. E., Jiao, H., Kim, S.-J., Yen, K., Mehta, H. H., Wan, J., Carpten, J. C. and Cohen, P. (2019). Comparing the Utility of Mitochondrial and Nuclear DNA to Adjust for Genetic Ancestry in Association Studies. *Cells* 8.4, 306–306. DOI: 10.3390/cells8040306.

Miller, B., Kim, S.-J., Kumagai, H., Mehta, H. H., Xiang, W., Liu, J., Yen, K. and Cohen, P. (2020). Peptides Derived from Small Mitochondrial Open Reading Frames: Genomic, Biological, and Therapeutic Implications. *Exp. Cell Res.* 393.2, 112056. DOI: 10.1016/j.yexcr.2020.112056.

Milot, E., Moreau, C., Gagnon, A., Cohen, A. A., Brais, B. and Labuda, D. (2017). Mother's Curse Neutralizes Natural Selection against a Human Genetic Disease over Three Centuries. *Nat. Ecol. Evol.* 1.9, 1400–1406. DOI: 10.1038/s41559-017-0276-6.

Min, J. L. et al. (2021). Genomic and Phenotypic Insights from an Atlas of Genetic Effects on DNA Methylation. *Nat. Genet.* 53.9, 1311–1321. DOI: 10.1038/s41588-021-00923-x.

Mort, M., Evani, U. S., Krishnan, V. G., Kamati, K. K., Baenziger, P. H., Bagchi, A., Peters, B. J., Sathyesh, R., Li, B., Sun, Y., Xue, B., Shah, N. H., Kann, M. G., Cooper, D. N., Radivojac, P. and Mooney, S. D. (2010). In Silico Functional Profiling of Human Disease-Associated and Polymorphic Amino Acid Substitutions. *Hum. Mutat.* 31.3, 335–346. DOI: 10.1002/humu.21192.

Mossman, J. A., Slate, J., Birkhead, T. R., Moore, H. D. and Pacey, A. A. (2012). Mitochondrial Haplotype Does Not Influence Sperm Motility in a UK Population of Men. *Hum. Reprod.* 27.3, 641–651. DOI: 10.1093/humrep/der438.

Mossman, J. A., Tross, J. G., Li, N., Wu, Z. and Rand, D. M. (2016). Mitochondrial-Nuclear Interactions Mediate Sex-Specific Transcriptional Profiles in Drosophila. *Genetics* 204.2, 613–630. DOI: 10.1534/genetics.116.192328.

Mottis, A., Herzig, S. and Auwerx, J. (2019). Mitocellular Communication: Shaping Health and Disease. *Science* 366.6467, 827–832. DOI: 10.1126/science.aax3768.

Mueller, E. E., Eder, W., Ebner, S., Schwaiger, E., Santic, D., Kreindl, T., Stanger, O., Paulweber, B., Iglseder, B., Oberkofler, H., Maier, R., Mayr, J. A., Krempler, F., Weitgasser, R., Patsch, W., Sperl, W., Kofler, B. and Federici, M. (2011). The Mitochondrial T16189C Polymorphism Is Associated with Coronary Artery Disease in Middle European Populations. *PLOS ONE* 6.1, e16455. DOI: 10.1371/journal.pone.0016455.

Muller, H. J. (1964). The Relation of Recombination to Mutational Advance. *Mut. Res. - Fund. Mol. M.* 1.1, 2–9. DOI: 10.1016/0027-5107(64)90047-8.

Nagarajan-Radha, V., Aitkenhead, I., Clancy, D. J., Chown, S. L. and Dowling, D. K. (2020). Sex-Specific Effects of Mitochondrial Haplotype on Metabolic Rate in Drosophila Melanogaster Support Predictions of the Mother's Curse Hypothesis. *Philos. Trans. R. Soc. B* 375.1790, 20190178. DOI: 10.1098/rstb.2019.0178.

Naue, J., Hörer, S., Sänger, T., Strobl, C., Hatzer-Grubwieser, P., Parson, W. and Lutz-Bonengel, S. (2015). Evidence for Frequent and Tissue-Specific Sequence Heteroplasmy in Human Mitochondrial DNA. *Mitochondrion* 20, 82–94. DOI: 10.1016/j.mito.2014.12.002.

Neckelmann, N., Li, K., Wade, R. P., Shuster, R. and Wallace, D. C. (1987). cDNA Sequence of a Human Skeletal Muscle ADP/ATP Translocator: Lack of a Leader Peptide, Divergence from a Fibroblast Translocator cDNA, and Coevolution with Mitochondrial DNA Genes. *Proc. Natl. Acad. Sci. U.S.A.* 84.21, 7580–7584. DOI: 10.1073/pnas.84.21.7580.

Nieminen, T., Lehtinen, R., Viik, J., Lehtimäki, T., Niemelä, K., Nikus, K., Niemi, M., Kallio, J., Kööbi, T., Turjanmaa, V. and Kähönen, M. (2006). The Finnish Cardiovascular Study (FINCAVAS): Characterising Patients with High Risk of Cardiovascular Morbidity and Mortality. *BMC Cardiovasc. Disord.* 6, 9. DOI: 10.1186/1471-2261-6-9.

Nissanka, N. and Moraes, C. T. (2018). Mitochondrial DNA Damage and Reactive Oxygen Species in Neurodegenerative Disease. *FEBS Lett.* 592.5, 728–742. DOI: 10.1002/1873-3468.12956.

Nolfi-Donegan, D., Braganza, A. and Shiva, S. (2020). Mitochondrial Electron Transport Chain: Oxidative Phosphorylation, Oxidant Production, and Methods of Measurement. *Redox Biol.* 37, 101674. DOI: 10.1016/j.redox.2020.101674.

Nyholt, D. R. (2004). A Simple Correction for Multiple Testing for Single-Nucleotide Polymorphisms in Linkage Disequilibrium with Each Other. *Am. J. Hum. Genet.* 74.4, 765–769. DOI: 10.1086/383251.

Okbay, A. et al. (2016). Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment. *Nature* 533.7604, 539–542. DOI: 10.1038/nature17671.

Orlanski, S., Labi, V., Reizel, Y., Spiro, A., Lichtenstein, M., Levin-Klein, R., Koralov, S. B., Skversky, Y., Rajewsky, K., Cedar, H. and Bergman, Y. (2016). Tissue-Specific DNA Demethylation Is Required for Proper B-cell Differentiation and Function. *Proc. Natl. Acad. Sci. U.S.A.* 113.18, 5018–5023. DOI: 10.1073/pnas.1604365113.

Oshiumi, H., Matsumoto, M., Hatakeyama, S. and Seya, T. (2009). Riplet/RNF135, a RING Finger Protein, Ubiquitinates RIG-I to Promote Interferon-Beta Induction during the Early Phase of Viral Infection. *J. Biol. Chem.* 284.2, 807–817. DOI: 10.1074/jbc.M804259200.

Pagnamenta, A. T., Wei, W., Rahman, S. and Chinnery, P. F. (2021). Biparental Inheritance of Mitochondrial DNA Revisited. *Nat. Rev. Genet.* 22.8, 477–478. DOI: 10.1038/s41576-021-00380-6.

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. *Hum. Mol. Genet.* 22.2, 384–390. DOI: 10.1093/hmg/dds435.

Pedersen, K., Haupt-Jorgensen, M., Krogvold, L., Kaur, S., Gerling, I. C., Pociot, F., Dahl-Joergensen, K. and Buschard, K. (2021). Genetic Predisposition in the 2'-5'A Pathway in the Development of Type 1 Diabetes: Potential Contribution to Dysregulation of Innate Antiviral Immunity. *Diabetologia* 64.8, 1805–1815. DOI: 10.1007/s00125-021-05469-5.

Pereira, L., Gonçalves, J., Franco-Duarte, R., Silva, J., Rocha, T., Arnold, C., Richards, M. and Macaulay, V. (2007). No Evidence for an mtDNA Role in Sperm Motility: Data from Complete Sequencing of Asthenozoospermic Males. *Mol. Biol. Evol.* 24.3, 868–874. DOI: 10.1093/molbev/msm004.

Pereira, L., Soares, P., Radivojac, P., Li, B. and Samuels, D. C. (2011). Comparing Phylogeny and the Predicted Pathogenicity of Protein Variations Reveals Equal Purifying Selection across the Global Human mtDNA Diversity. *Am. J. Hum. Genet.* 88.4, 433–439. DOI: 10.1016/J.AJHG.2011.03.006.

Pesole, G., Gissi, C., De Chirico, A. and Saccone, C. (1999). Nucleotide Substitution Rate of Mammalian Mitochondrial Genomes. *J. Mol. Evol.* 48.4, 427–434. DOI: 10.1007/PL00006487.

Pfanner, N., Warscheid, B. and Wiedemann, N. (2019). Mitochondrial Proteins: From Biogenesis to Functional Networks. *Nat. Rev. Mol. Cell Biol.* 20.5, 267–284. DOI: 10.1038/s41580-018-0092-0.

Price, A. L., Zaitlen, N. A., Reich, D. and Patterson, N. (2010). New Approaches to Population Stratification in Genome-Wide Association Studies. *Nat. Rev. Genet.* 11.7, 459–463. DOI: 10.1038/nrg2813.

Quirós, P. M., Mottis, A. and Auwerx, J. (2016). Mitonuclear Communication in Homeostasis and Stress. *Nat. Rev. Mol. Cell Biol.* 17.4, 213–226. DOI: 10.1038/nrm.2016.23.

Rabenstein, A., Catarino, C. B., Rampeltshammer, V., Schindler, D., Gallenmüller, C., Priglinger, C., Pogarell, O., Rüther, T. and Klopstock, T. (2021). Smoking and Alcohol, Health-Related Quality of Life and Psychiatric Comorbidities in Leber's Hereditary Optic Neuropathy Mutation Carriers: A Prospective Cohort Study. *Orphanet J. Rare Dis.* 16.1, 127. DOI: 10.1186/s13023-021-01724-5.

Raitakari, O. T., Juonala, M., Rönnemaa, T., Keltikangas-Järvinen, L., Räsänen, L., Pietikäinen, M., Hutri-Kähönen, N., Taittonen, L., Jokinen, E., Marniemi, J., Jula, A., Telama, R., Kähönen, M., Lehtimäki, T., Åkerblom, H. K. and Viikari, J. S. (2008). Cohort Profile: The Cardiovascular Risk in Young Finns Study. *Int. J. Epidemiol.* 37.6, 1220–1226. DOI: 10.1093/ije/dym225.

Ramanjaneya, M., Bettahi, I., Jerobin, J., Chandra, P., Abi Khalil, C., Skarulis, M., Atkin, S. L. and Abou-Samra, A.-B. (2019). Mitochondrial-Derived Peptides Are Down Regulated in Diabetes Subjects. *Front. Endocrinol.* 10, 331.

Reynolds, J. C., Lai, R. W., Woodhead, J. S. T., Joly, J. H., Mitchell, C. J., Cameron-Smith, D., Lu, R., Cohen, P., Graham, N. A., Benayoun, B. A., Merry, T. L. and Lee, C. (2021). MOTS-c Is an Exercise-Induced Mitochondrial-Encoded Regulator of Age-Dependent Physical Decline and Muscle Homeostasis. *Nat. Commun.* 12.1, 470. DOI: `10.1038/s41467-020-20790-0`.

Ricchetti, M., Tekaia, F. and Dujon, B. (2004). Continued Colonization of the Human Genome by Mitochondrial DNA. *PLOS Biol.* 2.9, e273. DOI: `10.1371/journal.pbio.0020273`.

Ricquier, D. and Bouillaud, F. (2000). Mitochondrial Uncoupling Proteins: From Mitochondria to the Regulation of Energy Balance. *J. Physiol.* 529.1, 3–10. DOI: `10.1111/j.1469-7793.2000.00003.x`.

Ristow, M. and Schmeisser, K. (2014). Mitohormesis: Promoting Health and Lifespan by Increased Levels of Reactive Oxygen Species (ROS). *Dose-Response* 12.2, 288–341. DOI: `10.2203/dose-response.13-035.Ristow`.

Ristow, M., Zarse, K., Oberbach, A., Klöting, N., Birringer, M., Kiehntopf, M., Stumvoll, M., Kahn, C. R. and Blüher, M. (2009). Antioxidants Prevent Health-Promoting Effects of Physical Exercise in Humans. *Proc. Natl. Acad. Sci. U.S.A.* 106.21, 8665–8670. DOI: `10.1073/pnas.0903485106`.

Rius, R., Cowley, M. J., Riley, L., Puttick, C., Thorburn, D. R. and Christodoulou, J. (2019). Biparental Inheritance of Mitochondrial DNA in Humans Is Not a Common Phenomenon. *Genet. Med.* 21.12, 2823–2826. DOI: `10.1038/s41436-019-0568-0`.

Rizzuto, R., De Stefani, D., Raffaello, A. and Mammucari, C. (2012). Mitochondria as Sensors and Regulators of Calcium Signalling. *Nat. Rev. Mol. Cell Biol.* 13.9, 566–578. DOI: `10.1038/nrm3412`.

Rorbach, J. and Minczuk, M. (2012). The Post-Transcriptional Life of Mammalian Mitochondrial RNA. *Biochem. J.* 444.3, 357–373. DOI: `10.1042/BJ20112208`.

Ruiz-Pesini, E., Lapeña, A.-C., Díez-Sánchez, C., Pérez-Martos, A., Montoya, J., Alvarez, E., Díaz, M., Urriés, A., Montoro, L., López-Pérez, M. J. and Enríquez, J. A. (2000). Human mtDNA Haplogroups Associated with High or Reduced Spermatozoa Motility. *Am. J. Hum. Genet.* 67.3, 682–696. DOI: `10.1086/303040`.

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, 223–226. DOI: `10.1126/science.1088434`.

Saha, S. K., Saba, A. A., Hasib, M., Rimon, R. A., Hasan, I., Alam, M. S., Mahmud, I. and Nabi, A. H. M. N. (2021). Evaluation of D-loop Hypervariable Region I Variations, Haplogroups and Copy Number of Mitochondrial DNA in Bangladeshi Population with Type 2 Diabetes. *Heliyon* 7.7, e07573. DOI: 10.1016/j.heliyon.2021.e07573.

Salas, A., García-Magariños, M., Logan, I. and Bandelt, H. J. (2014). The Saga of the Many Studies Wrongly Associating Mitochondrial DNA with Breast Cancer. *BMC Cancer* 14, 659. DOI: 10.1186/1471-2407-14-659.

Samuels, D. C., Li, C., Li, B., Song, Z., Torstenson, E., Clay, H. B., 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, e1003929. DOI: 10.1371/journal.pgen.1003929.

Santoro, A. et al. (2010). Evidence for Sub-Haplogroup H5 of Mitochondrial DNA as a Risk Factor for Late Onset Alzheimer's Disease. *PLOS ONE* 5.8, e12037. DOI: 10.1371/journal.pone.0012037.

Saxena, R., De Bakker, P. I. W., 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, 54–61. DOI: 10.1086/504926.

Schon, E. A., DiMauro, S. and Hirano, M. (2012). Human Mitochondrial DNA: Roles of Inherited and Somatic Mutations. *Nat. Rev. Genet.* 13.12, 878–890. DOI: 10.1038/nrg3275.

Schwartz, M. and Vissing, J. (2002). Paternal Inheritance of Mitochondrial DNA. *N. Engl. J. Med.* 347.8, 576–580. DOI: 10.1056/NEJMoa020350.

Shabalin, A. A. (2012). Matrix eQTL: Ultra Fast eQTL Analysis via Large Matrix Operations. *Bioinformatics* 28.10, 1353–1358. DOI: 10.1093/bioinformatics/bts163.

Shadel, G. S. and Horvath, T. L. (2015). Mitochondrial ROS Signaling in Organismal Homeostasis. *Cell* 163.3, 560–569. DOI: 10.1016/j.cell.2015.10.001.

Singmann, P. et al. (2015). Characterization of Whole-Genome Autosomal Differences of DNA Methylation between Men and Women. *Epigenetics Chromatin* 8, 43. DOI: 10.1186/s13072-015-0035-3.

Smiraglia, D., Kulawiec, M., Bistulfi, G. L., Ghoshal, S. and Singh, K. K. (2008). A Novel Role for Mitochondria in Regulating Epigenetic Modifications in the Nucleus. *Cancer Biol. Ther.* 7.8, 1182–1190. DOI: 10.4161/cbt.7.8.6215.

Sonney, S., Leipzig, J., Lott, M. T., Zhang, S., Procaccio, V., Wallace, D. C. and Sondheimer, N. (2017). Predicting the Pathogenicity of Novel Variants in Mitochondrial tRNA with MitoTIP. *PLOS Comp. Biol.* 13.12, e1005867. DOI: 10.1371/journal.pcbi.1005867.

Steves, C. J., Spector, T. D. and Jackson, S. H. D. (2012). Ageing, Genes, Environment and Epigenetics: What Twin Studies Tell Us Now, and in the Future. *Age Ageing* 41.5, 581–586. DOI: 10.1093/ageing/afs097.

Stewart, J. B. and Chinnery, P. F. (2015). The Dynamics of Mitochondrial DNA Heteroplasmy: Implications for Human Health and Disease. *Nat. Rev. Genet.* 16.9, 530–542. DOI: 10.1038/nrg3966.

Stewart, J. B. and Chinnery, P. F. (2021). Extreme Heterogeneity of Human Mitochondrial DNA from Organelles to Populations. *Nat. Rev. Genet.* 22.2, 1–13. DOI: 10.1038/s41576-020-00284-x.

Stewart, J. B., Freyer, C., Elson, J. L. and Larsson, N.-G. (2008). Purifying Selection of mtDNA and Its Implications for Understanding Evolution and Mitochondrial Disease. *Nat. Rev. Genet.* 9.9, 657–662. DOI: 10.1038/nrg2396.

Sulonen, A.-M., Ellonen, P., Almusa, H., Lepistö, M., Eldfors, S., Hannula, S., Miettinen, T., Tyynismaa, H., Salo, P., Heckman, C., Joensuu, H., Raivio, T., Suomalainen, A. and Saarela, J. (2011). Comparison of Solution-Based Exome Capture Methods for next Generation Sequencing. *Genome Biol.* 12.9, R94. DOI: 10.1186/gb-2011-12-9-r94.

Suomalainen, A. (2019). Mitochondrial DNA Inheritance in Humans: Mix, Match, and Survival of the Fittest. *Cell Metab.* 30.2, 231–232. DOI: 10.1016/j.cmet.2019.07.009.

Tabák, A. G., Herder, C., Rathmann, W., Brunner, E. J. and Kivimäki, M. (2012). Prediabetes: A High-Risk State for Diabetes Development. *The Lancet* 379, 2279–2290. DOI: 10.1016/S0140.

Takasaki, S. (2009). Mitochondrial Haplogroups Associated with Japanese Centenarians, Alzheimer's Patients, Parkinson's Patients, Type 2 Diabetic Patients and Healthy Non-Obese Young Males. *J. Genet. Genom.* 36.7, 425–434. DOI: 10.1016/S1673-8527(08)60132-0.

Tamura, K. and Nei, M. (1993). Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees. *Mol. Biol. Evol.* 10.3, 512–526. DOI: 10.1093/oxfordjournals.molbev.a040023.

Tang, Z. Z. and Lin, D. Y. (2015). Meta-Analysis for Discovering Rare-Variant Associations: Statistical Methods and Software Programs. *Am. J. Hum. Genet.* 97.1, 35–53. DOI: 10.1016/j.ajhg.2015.05.001.

Taylor, R. W., McDonnell, M. T., Blakely, E. L., Chinnery, P. F., Taylor, G. A., Howell, N., Zeviani, M., Briem, E., Carrara, F. and Turnbull, D. M. (2003). Genotypes from Patients Indicate No Paternal Mitochondrial DNA Contribution. *Ann. Neurol.* 54.4, 521–524. DOI: 10.1002/ana.10673.

Teng, L., Zheng, J., Leng, J. and Ding, Y. (2012). Clinical and Molecular Characterization of a Han Chinese Family with High Penetrance of Essential Hypertension. *Mitochondrial DNA* 23.6, 461–465. DOI: 10.3109/19401736.2012.710205.

Tobin, M. D., Sheehan, N. A., Scurrah, K. J. and Burton, P. R. (2005). Adjusting for Treatment Effects in Studies of Quantitative Traits: Antihypertensive Therapy and Systolic Blood Pressure. *Stat. Med.* 24.19, 2911–2935. DOI: 10.1002/sim.2165.

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, 1835–1850. DOI: 10.1093/genetics/144.4.1835.

Tragante, V. et al. (2014). Gene-Centric Meta-analysis in 87,736 Individuals of European Ancestry Identifies Multiple Blood-Pressure-Related Loci. *Am. J. Hum. Genet.* 94.3, 349–360. DOI: 10.1016/j.ajhg.2013.12.016.

Tretter, L. and Adam-Vizi, V. (2005). Alpha-Ketoglutarate Dehydrogenase: A Target and Generator of Oxidative Stress. *Philos. Trans. R. Soc. B* 360.1464, 2335–2345. DOI: 10.1098/rstb.2005.1764.

Van Dongen, J., Nivard, M. G., Willemsen, G., Hottenga, J.-J., Helmer, Q., Dolan, C. V., Ehli, E. A., Davies, G. E., van Iterson, M., Breeze, C. E., Beck, S., Suchiman, H. E., Jansen, R., van Meurs, J. B., Heijmans, B. T., Slagboom, P. E. and Boomsma, D. I. (2016). Genetic and Environmental Influences Interact with Age and Sex in Shaping the Human Methylome. *Nat. Commun.* 7.1, 1–13. DOI: 10.1038/ncomms11115.

Van Iterson, M., van Zwet, E. W., Heijmans, B. T. and the BIOS Consortium (2017). Controlling Bias and Inflation in Epigenome- and Transcriptome-Wide Association Studies Using the Empirical Null Distribution. *Genome Biol.* 18, 19. DOI: 10.1186/s13059-016-1131-9.

Van Oven, M. (2015). PhyloTree Build 17: Growing the Human Mitochondrial DNA Tree. *Forensic Sci. Int. Genet. Suppl. Ser.* 5, e392–e394. DOI: 10.1016/J.FSIGSS.2015.09.155.

Van Oven, M. and Kayser, M. (2009). Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation. *Hum. Mutat.* 30.2, E386–E394. DOI: 10.1002/humu.20921.

Vázquez-Acevedo, M., Antaramian, A., Corona, N. and González-Halphen, D. (1993). Subunit Structures of Purified Beef Mitochondrial Cytochromebc1 Complex from Liver and Heart. *J. Bioenerg. Biomembr.* 25.4, 401–410. DOI: 10.1007/BF00762466.

Venter, M., Malan, L., van Dyk, E., Elson, J. L. and van der Westhuizen, F. H. (2017). Using MutPred Derived mtDNA Load Scores to Evaluate mtDNA Variation in Hypertension and Diabetes in a Two-Population Cohort: The SABPA Study. *J. Genet. Genomics* 44.3, 139–149. DOI: 10.1016/J.JGG.2016.12.003.

Vivian, C. J., Brinker, A. E., Graw, S., Koestler, D. C., Legendre, C., Gooden, G. C., Salhia, B. and Welch, D. R. (2017). Mitochondrial Genomic Backgrounds Affect Nuclear DNA Methylation and Gene Expression. *Cancer Res.* 77.22, 6202–6214. DOI: 10.1158/0008-5472.CAN-17-1473.

Vona, B. et al. (2018). Expanding the Clinical Phenotype of IARS2-related Mitochondrial Disease. *BMC Med. Genet.* 19, 196. DOI: 10.1186/s12881-018-0709-3.

Walaszczyk, E., Luijten, M., Spijkerman, A. M. W., Bonder, M. J., Lutgers, H. L., Snieder, H., Wolffenbuttel, B. H. R. and van Vliet-Ostaptchouk, J. V. (2018). DNA Methylation Markers Associated with Type 2 Diabetes, Fasting Glucose and HbA 1c Levels: A Systematic Review and Replication in a Case-Control Sample of the Lifelines Study. *Diabetologia* 61.2, 354–368. DOI: 10.1007/s00125-017-4497-7.

Wallace, D. C. (2007). Why Do We Still Have a Maternally Inherited Mitochondrial DNA? Insights from Evolutionary Medicine. *Annu. Rev. Biochem.* 76, 781–821. DOI: 10.1146/annurev.biochem.76.081205.150955.

Wallace, D. C. (2010). Mitochondrial DNA Mutations in Disease and Aging. *Environ. Mol. Mutagen.* 51.5, 440–450. DOI: 10.1002/em.20586.

Wallace, D. C., Brown, M. D. and Lott, M. T. (1999). Mitochondrial DNA Variation in Human Evolution and Disease. *Gene* 238.1, 211–230. DOI: 10.1016/S0378-1119(99)00295-4.

Wang, J., Zhao, Q., Hastie, T. and Owen, A. B. (2017). Confounder Adjustment in Multiple Hypothesis Testing. *Ann. Stat.* 45.5, 1863–1894. DOI: 10.1214/16-AOS1511.

Wang, P. et al. (2022). Epigenome-Wide Association Study of Mitochondrial Genome Copy Number. *Hum. Mol. Genet.* 31.2, 309–319. DOI: 10.1093/hmg/ddab240.

Wang, S., Li, R., Fettermann, A., Li, Z., Qian, Y., Liu, Y., Wang, X., Zhou, A., Mo, J. Q., Yang, L., Jiang, P., Taschner, A., Rossmanith, W. and Guan, M.-X. (2011). Maternally Inherited Essential Hypertension Is Associated With the Novel 4263A>G Mutation in the Mitochondrial tRNA Ile Gene in a Large Han Chinese Family. *Circ. Res.* 108.7, 862–870. DOI: 10.1161/CIRCRESAHA.110.231811.

Wang, X., Kang, D. D., Shen, K., Song, C., Lu, S., Chang, L.-C., Liao, S. G., Huo, Z., Tang, S. and Ding, Y. (2012). An R Package Suite for Microarray Meta-Analysis in Quality Control, Differentially Expressed Gene Analysis and Pathway Enrichment Detection. *Bioinformatics* 28.19, 2534–2536. DOI: 10.1093/bioinformatics/bts485.

Wei, W. and Chinnery, P. F. (2020). Inheritance of Mitochondrial DNA in Humans: Implications for Rare and Common Diseases. *J. Intern. Med.* 287, 634–644. DOI: 10.1111/joim.13047.

Wei, W. et al. (2019). Germline Selection Shapes Human Mitochondrial DNA Diversity. *Science* 364.6442, eaau6520. DOI: 10.1126/science.aau6520.

Weinberg, S. E., Singer, B. D., Steinert, E. M., Martinez, C. A., Mehta, M. M., Martínez-Reyes, I., Gao, P., Helmin, K. A., Abdala-Valencia, H., Sena, L. A., Schumacker, P. T., Turka, L. A. and Chandel, N. S. (2019). Mitochondrial Complex III Is Essential for Suppressive Function of Regulatory T Cells. *Nature* 565.7740, 495–499. DOI: 10.1038/s41586-018-0846-z.

Weissensteiner, H., Forer, L., Fendt, L., Kheirkhah, A., Salas, A., Kronenberg, F. and Schoenherr, S. (2021). Contamination Detection in Sequencing Studies Using the Mitochondrial Phylogeny. *Genome Res.* 31.2, 309–316. DOI: 10.1101/gr.256545.119.

Weissensteiner, H., Forer, L., Fuchsberger, C., Schöpf, B., Kloss-Brandstätter, A., Specht, G., Kronenberg, F. and Schönherr, S. (2016a). mtDNA-Server: Next-Generation

Sequencing Data Analysis of Human Mitochondrial DNA in the Cloud. *Nucleic Acids Res.* 44.W1, W64–W69. DOI: 10.1093/nar/gkw247.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A. and Schönherr, S. (2016b). HaploGrep 2: Mitochondrial Haplogroup Classification in the Era of High-Throughput Sequencing. *Nucleic Acids Res.* 44.W1, W58–W63. DOI: 10.1093/nar/gkw233.

West, A. P. and Shadel, G. S. (2017). Mitochondrial DNA in Innate Immune Responses and Inflammatory Pathology. *Nat. Rev. Immunol.* 17.6, 363–375. DOI: 10.1038/nri.2017.21.

Winkelmann, B. R., März, W., Boehm, B. O., Zotz, R., Hager, J., Hellstern, P. and Senges, J. (2001). Rationale and Design of the LURIC Study - a Resource for Functional Genomics, Pharmacogenomics and Long-Term Prognosis of Cardiovascular Disease. *Pharmacogenomics* 2, S1–S73. DOI: 10.1517/14622416.2.1.S1.

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M. and Lin, X. (2011). Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. *Am. J. Hum. Genet.* 89.1, 82–93. DOI: 10.1016/J.AJHG.2011.05.029.

Yang, J. et al. (2011). Genomic Inflation Factors under Polygenic Inheritance. *Eur. J. Hum. Genet.* 19.7, 807–812. DOI: 10.1038/ejhg.2011.39.

Yang, S. Y., Castellani, C. A., Longchamps, R. J., Pillalamarri, V. K., O'Rourke, B., Guallar, E. and Arking, D. E. (2021). Blood-Derived Mitochondrial DNA Copy Number Is Associated with Gene Expression across Multiple Tissues and Is Predictive for Incident Neurodegenerative Disease. *Genome Res.* 31.3, 349–358. DOI: 10.1101/gr.269381.120.

Zarse, K. and Ristow, M. (2015). A Mitochondrially Encoded Hormone Ameliorates Obesity and Insulin Resistance. *Cell Metab.* 21.3, 355–356. DOI: 10.1016/j.cmet.2015.02.013.

Zhang, P., Samuels, D. C., Wang, J., Zhao, S., Shyr, Y. and Guo, Y. (2016). Mitochondria Single Nucleotide Variation across Six Blood Cell Types. *Mitochondrion* 28, 16–22. DOI: 10.1016/j.mito.2016.03.001.

Zhang, W., Qu, J., Liu, G.-H. and Belmonte, J. C. I. (2020). The Ageing Epigenome and Its Rejuvenation. *Nat. Rev. Mol. Cell Biol.* 21.3, 137–150. DOI: 10.1038/s41580-019-0204-5.

Zhang, D.-X. and Hewitt, G. M. (1996). Nuclear Integrations: Challenges for Mitochondrial DNA Markers. *Trends Ecol. Evol.* 11.6, 247–251. DOI: 10.1016/0169-5347(96)10031-8.

Zhao, R.-Z., Jiang, S., Zhang, L. and Yu, Z.-B. (2019). Mitochondrial Electron Transport Chain, ROS Generation and Uncoupling. *Int. J. Mol. Med.* 44.1, 3–15. DOI: 10.3892/ijmm.2019.4188.

ORIGINAL PUBLICATIONS

PUBLICATION

I

Discovery of mitochondrial DNA variants associated with genome-wide blood cell gene expression: a population-based mtDNA sequencing study

Laaksonen, J., Seppälä, I., Raitoharju, E., Mononen, N., Lyytikäinen, L.-P., Waldenberger, M., Illig, T., Lepistö, M., Almusa, H., Ellonen, P., Hutri-Kähönen, N., Juonala, M., Kähönen, M., Raitakari, O., Salonen, J. T. and Lehtimäki, T.

Human Molecular Genetics 28.8 (2019), 1381–1391

Publication reprinted with the permission of the copyright holders

Human Molecular Genetics, 2019, Vol. 28, No. 8 1381–1391

doi: 10.1093/hmg/ddz011
Advance Access Publication Date: 9 January 2019
Association Studies Article

OXFORD

ASSOCIATION STUDIES ARTICLE

Discovery of mitochondrial DNA variants associated with genome-wide blood cell gene expression: a population-based mtDNA sequencing study

Jaakko Laaksonen[1,†,*], Ilkka Seppälä[1], Emma Raitoharju[1], Nina Mononen[1], Leo-Pekka Lyytikäinen[1], Melanie Waldenberger[2,3], Thomas Illig[4,5], Maija Lepistö[6], Henrikki Almusa[6], Pekka Ellonen[6], Nina Hutri-Kähönen[7], Markus Juonala[8,9,10], Mika Kähönen[11], Olli Raitakari[12,13], Jukka T. Salonen[14,15] and Terho Lehtimäki[1]

[1]Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, 33520, Finland [2]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany [3]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany [4]Hannover Unified Biobank, Hannover Medical School, Hannover, 30625, Germany [5]Institute for Human Genetics, Hannover Medical School, Hannover, 30625, Germany [6]Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, 00290, Finland [7]Department of Pediatrics, Tampere University Hospital and Faculty of Medicine and Health Technology, Tampere University, Tampere, 33520, Finland [8]Department of Medicine, University of Turku, Turku, 20520, Finland [9]Division of Medicine, Turku University Hospital, Turku, 20520, Finland [10]Murdoch Children's Research Institute, Parkville, Victoria, 3052, Australia [11]Department of Clinical Physiology, Tampere University Hospital and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, 33520, Finland [12]Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, 20520, Finland [13]Research Centre for Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, 20520, Finland [14]Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, 00014, Finland and [15]MAS-Metabolic Analytical Services Oy, Helsinki, 00990, Finland

*To whom correspondence should be addressed at: Faculty of Medicine and Health Technology, PO Box 100, FI-33014 Tampere University, Finland. Tel: +358 504080774; Fax: +358 32599003; Email: jaakko.h.laaksonen@tuni.fi

Downloaded from https://academic.oup.com/hmg/article/28/8/1381/5281396 by guest on 16 November 2020

Abstract

The effect of mitochondrial DNA (mtDNA) variation on peripheral blood transcriptomics in health and disease is not fully known. Sex-specific mitochondrially controlled gene expression patterns have been shown in *Drosophila melanogaster* but in humans, evidence is lacking. Functional variation in mtDNA may also have a role in the development of type 2 diabetes and its precursor state, i.e. prediabetes. We examined the associations between mitochondrial single-nucleotide polymorphisms

[†]Jaakko Laaksonen, http://orcid.org/0000-0003-3577-6928
Received: October 16, 2018. **Revised:** December 14, 2018. **Accepted:** January 7, 2019

© The Author(s) 2019. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com

(mtSNPs) and peripheral blood transcriptomics with a focus on sex- and prediabetes-specific effects. The genome-wide blood cell expression data of 19 637 probes, 199 deep-sequenced mtSNPs and nine haplogroups of 955 individuals from a population-based Young Finns Study cohort were used. Significant associations were identified with linear regression and analysis of covariance. The effects of sex and prediabetes on the associations between gene expression and mtSNPs were studied using random-effect meta-analysis. Our analysis identified 53 significant expression probe-mtSNP associations after Bonferroni correction, involving 7 genes and 31 mtSNPs. Eight probe-mtSNP signals remained independent after conditional analysis. In addition, five genes showed differential expression between haplogroups. The meta-analysis did not show any significant differences in linear model effect sizes between males and females but identified the association between the *OASL* gene and mtSNP C16294T to show prediabetes-specific effects. This study pinpoints new independent mtSNPs associated with peripheral blood transcriptomics and replicates six previously reported associations, providing further evidence of the mitochondrial genetic control of blood cell gene expression. In addition, we present evidence that prediabetes might lead to perturbations in mitochondrial control.

Downloaded from https://academic.oup.com/hmg/article/28/8/1381/5281396 by guest on 16 November 2020

Introduction

Mitochondrial DNA (mtDNA) is a maternally inherited, circular molecule containing ∼16 600 nucleotides that encode 22 transfer RNAs, 2 ribosomal RNAs and 13 polypeptides (1). It has a high mutation rate and these polymorphisms have accumulated during evolution, dividing the human population into smaller mitochondrial haplogroups. The definition of these haplogroups is based on particular combinations of certain single-nucleotide polymorphisms (SNPs) in mtDNA. Mitochondrial haplogroups are also associated with geographic areas and populations. For example, virtually all Scandinavian mtDNA falls into 10 different haplogroups: H, I, J, K, M, T, U, V, W and X (2,3).

mtDNA is transmitted mainly via the maternal lineage, which could create a male–female asymmetry in the expected severity of mitochondrial disease (4). The mitochondrial SNPs (mtSNPs) that are deleterious to males but not to females, such as those that impair sperm function, will not be subject to natural selection. This may play an important role for male-specific effects in health and disease (5). This hypothesis was tested in a study conducted with *Drosophila melanogaster*, which found sex-specific asymmetry in nuclear gene expression patterns. A strong effect of mtSNPs on nuclear gene expression was only observed in males, in females the mitochondrial effect was negligible (6). A non-sequencing-based study conducted on humans showed 15 significant associations between mtSNPs and nuclear gene expression but found little evidence of a sex-specific mitochondrial control of gene expression (7). The number of studied mtSNPs was only 78, and it is possible that the sex-specific effects are mediated via other mtSNPs not included in the study sample.

Interaction among mtDNA and nuclear DNA encoded factors is important to cellular function and in adaptation to environmental changes. Since the co-evolution of these two genomes has been demonstrated at a protein–protein interaction level and some modes of co-regulation at a transcriptional level also exist, it is logical that disruption in mitochondria–nuclear signaling may lead to disease (8,9). Some mtSNPs have been reported to be associated with type 2 diabetes (T2D), but this association has been seen only in Asians (10–12). A Korean study found mitochondrial haplogroups to be associated with an increased or decreased risk of T2D and with altered nuclear gene expression patterns that correlate with the susceptibility to develop T2D (13). However, it is unclear to what extent the mitochondrial–nuclear interaction is altered when the glucose homeostasis has already been impaired. Therefore, in addition to sex-specific differences of special interest in this study is prediabetes, a precursor state and a major risk factor for the development of T2D (14).

In this study, we wanted to (a) replicate the earlier gene expression–mtSNP associations (7), (b) to find new functional associations with a greater number of mtSNPs by using population-based mtDNA sequencing, (c) to study the differential gene expression between the major Scandinavian haplogroups present in our study population, (d) to investigate whether any gene expression–mtSNP associations show sex-specific differences and, finally, (e) to see whether gene expression–mtSNP associations are affected by the onset of prediabetes.

Results

Effect of mtSNPs and haplogroups on peripheral blood gene expression

A total of 3 907 763 expression probe-mtSNP pairs were tested for association in a linear regression model. The genomic inflation factor was 1.00 (quantile–quantile plot for expected versus observed *P*-values shown in Supplementary Material, Fig. S1), indicating that an inflation of the genetic association due to population stratification or undetected systematic error is unlikely. As shown in Table 1, a total of 53 expression probe-mtSNP pairs were significant after Bonferroni correction, corresponding to 5 nuclear and 2 mitochondrial genes and 31 mtSNPs. These seven identified genes regulated by mtSNPs were signal peptidase complex subunit 2 pseudogene 4 (*SPCS2P4*), ring finger protein 113A (*RNF113A*), signal peptidase complex subunit 2 (*SPCS2*), mitochondrially encoded cytochrome c oxidase II (*MT-CO2*), cardiolipin synthase 1 (*CRLS1*), solute carrier family 25 member 15 (*SLC25A14*) and mitochondrially encoded 16S RNA-like 1 (*MT-RNR2L1*).

The pairwise conditional analysis results for all significant mtSNPs with $r^2 > 0.30$ are shown in online Supplementary Material, Table S1. As expected, the results were influenced by the extent of pairwise correlations and the strength of individual associations. Five probe-mtSNP pairs that survived all pairwise conditional analyses are shown in Table 2. The correlation was very low between the two mtSNPs that associated with *MT-CO2* ($r^2 = 8.20 \times 10^{-4}$) and only one mtSNP associated with *MT-RNR2L1*; these associations were not subjected to conditional analysis. In total, six mtSNPs (G8269A, G9055A, A11251G, C15452T, A16162G and C16256T) had independent effect on gene expression. Figure 1 illustrates the normalized expression intensities for the top four genes relative to the alleles of the top associated independent mtSNP.

Analysis of covariance (ANCOVA) and Tukey's post hoc test indicated that five expression probes, corresponding to four nuclear and one mitochondrial genes, showed differential

Table 1. The 53 probe–mtSNP associations that had a p of $<1.28 \times 10^{-8}$

Gene	Illumina array ID	mtSNP	MAF	Beta	SE	P-value	Replication
SPCS2P4	4210315	G9055A	0.04	−1.11	0.08	1.81×10^{-39}	
SPCS2P4	4210315	A3480G	0.04	−1.09	0.09	4.84×10^{-32}	*
SPCS2P4	4210315	A10550G	0.04	−1.08	0.09	9.49×10^{-32}	*
SPCS2P4	4210315	C14167T	0.04	−1.08	0.09	9.49×10^{-32}	
SPCS2P4	4210315	T16224C	0.05	−0.95	0.08	1.77×10^{-30}	
SPCS2P4	4210315	T11299C	0.04	−1.03	0.09	5.94×10^{-29}	
SPCS2P4	4210315	T1189C	0.03	−1.04	0.09	1.17×10^{-27}	
RNF113A	6940196	G9055A	0.04	−1.23	0.13	2.48×10^{-21}	
SPCS2P4	4210315	T9698C	0.05	−0.74	0.08	1.01×10^{-19}	*
RNF113A	6940196	C14167T	0.04	−1.25	0.14	1.87×10^{-18}	
RNF113A	6940196	A10550G	0.04	−1.24	0.14	1.95×10^{-18}	
RNF113A	6940196	A3480G	0.04	−1.24	0.14	2.70×10^{-18}	
SPCS2P4	4210315	A9093G	0.02	−1.03	0.12	4.72×10^{-18}	
RNF113A	6940196	T11299C	0.04	−1.20	0.14	1.21×10^{-17}	
RNF113A	6940196	T1189C	0.03	−1.23	0.14	4.42×10^{-17}	
SPCS2P4	4210315	T9903C	0.02	−1.06	0.12	7.24×10^{-17}	
SPCS2P4	4210315	T14798C	0.10	−0.60	0.07	9.74×10^{-16}	
SPCS2P4	4210315	A1811G	0.08	−0.50	0.07	4.86×10^{-14}	
RNF113A	6940196	T16224C	0.05	−0.96	0.13	8.58×10^{-14}	
SPCS2	7040068	G9055A	0.04	−0.93	0.12	9.59×10^{-14}	
CRLS1	2710446	A8869G	0.02	−0.75	0.10	5.30×10^{-13}	
CRLS1	2710446	T4639C	0.02	−0.75	0.10	8.45×10^{-13}	
SPCS2P4	4210315	G11377A	0.03	−0.76	0.11	2.19×10^{-12}	
RNF113A	6940196	T9698C	0.05	−0.87	0.12	2.57×10^{-12}	
CRLS1	2710446	C5263T	0.02	−0.75	0.11	5.28×10^{-12}	
RNF113A	6940196	A9093G	0.02	−1.23	0.18	7.40×10^{-12}	
MT-CO2	6550386	G8269A	0.01	−1.62	0.24	2.09×10^{-11}	*
SPCS2P4	4210315	A11251G	0.12	1.04	0.15	2.64×10^{-11}	
SPCS2P4	4210315	C15452A	0.12	1.03	0.15	5.58×10^{-11}	
RNF113A	6940196	T9903C	0.02	−1.24	0.19	2.04×10^{-10}	
SLC25A14	1710754	A3505G	0.05	0.77	0.12	3.28×10^{-10}	
SLC25A14	1710754	T1243C	0.05	0.76	0.12	4.54×10^{-10}	
RNF113A	6940196	T14798C	0.10	−0.72	0.11	4.76×10^{-10}	
SPCS2	7040068	A3480G	0.04	−0.85	0.14	4.81×10^{-10}	*
SPCS2	7040068	A10550G	0.04	−0.85	0.14	5.04×10^{-10}	*
MT-RNR2L1	1230164	C16256T	0.07	0.39	0.06	5.28×10^{-10}	
SPCS2	7040068	C14167T	0.04	−0.85	0.14	5.44×10^{-10}	
SPCS2	7040068	T16224C	0.05	−0.76	0.12	9.72×10^{-10}	
SLC25A14	1710754	A11947G	0.05	0.75	0.12	1.03×10^{-9}	
SLC25A14	1710754	G8994A	0.05	0.74	0.12	1.09×10^{-9}	
RNF113A	6940196	T4216C	0.12	1.32	0.22	1.44×10^{-9}	
SLC25A14	1710754	G5046A	0.05	0.74	0.12	1.76×10^{-9}	
SPCS2P4	4210315	A10398G	0.14	−0.40	0.07	1.80×10^{-9}	
SLC25A14	1710754	G15884C	0.05	0.74	0.12	2.25×10^{-9}	
RNF113A	6940196	A1811G	0.08	−0.60	0.10	3.43×10^{-9}	
RNF113A	6940196	C15452A	0.12	1.39	0.24	5.22×10^{-9}	
MT-CO2	6550386	A16162G	0.06	−0.66	0.11	5.48×10^{-9}	
SPCS2	7040068	T11299C	0.04	−0.79	0.13	7.40×10^{-9}	
RNF113A	6940196	A11251G	0.12	1.36	0.23	7.84×10^{-9}	
SLC25A14	1710754	T12414C	0.05	0.69	0.12	8.01×10^{-9}	
SPCS2	7040068	T1189C	0.03	−0.82	0.14	8.27×10^{-9}	
SLC25A14	1710754	G5460A	0.05	0.65	0.11	9.93×10^{-9}	
RNF113A	6940196	G11377A	0.03	−0.93	0.16	1.18×10^{-8}	

Abbreviation: SD, standard deviance

The beta-coefficient represents the proportion of one SD change in normalized gene expression intensity (mean = 0, SD = 1). An asterisk (*) marks the associations that were replicated from the previous results (7).

expression between haplogroups ($\lambda_{GC} = 1.04$, Supplementary Material, Fig. S2). Three of the genes, *SPCS2P4, RNF113A* and *SLC25A14,* were also associated with individual mtSNPs, but the other two genes, solute carrier family 2 member 8 (*SLC2A8,* Illumina Array Address 5870326) and mitochondrially encoded NADH dehydrogenase 5 (*MT-ND5,* Illumina Array Address 4880477), were not identified in the probe-mtSNP analysis. The results of Tukey's post hoc test are shown in Table 3. Figure 2 illustrates the expression levels of the top two genes, *SPCS2P4* and *RNF113A*, which were significantly lower in haplogroup K

Downloaded from https://academic.oup.com/hmg/article/28/8/1381/5281396 by guest on 16 November 2020

Table 2. Independent mtSNP-gene association signals from the pairwise conditional analysis

Gene	mtSNP	Conditional *P*-value	Conditional beta	Conditional SE	mtSNP-specific Bonferroni-corrected *P*-value
SPCS2P4	G9055A	$1.80 \times 10^{-9} - 4.14 \times 10^{-29}$	−0.97 – −1.36	0.09–0.20	4.17×10^{-3}
SPCS2P4	A11251G	1.41×10^{-3}	0.97	0.30	5.00×10^{-2}
SPCS2P4	C15452A	4.28×10^{-6}	1.13	0.24	5.00×10^{-2}
RNF113A	G9055A	$1.40 \times 10^{-4} - 2.82 \times 10^{-14}$	−1.18 – −1.36	0.14–0.32	4.17×10^{-3}
SPCS2	G9055A	$3.68 \times 10^{-5} - 1.38 \times 10^{-6}$	−0.90 – −1.39	0.21–0.31	8.33×10^{-3}

The beta-coefficient represents the proportion of one SD change in normalized gene expression intensity (mean = 0, SD = 1). The mtSNP-specific Bonferroni-corrected *P*-value accounts for the number of pairwise analyses made for each mtSNP and is defined as the limit of significance.

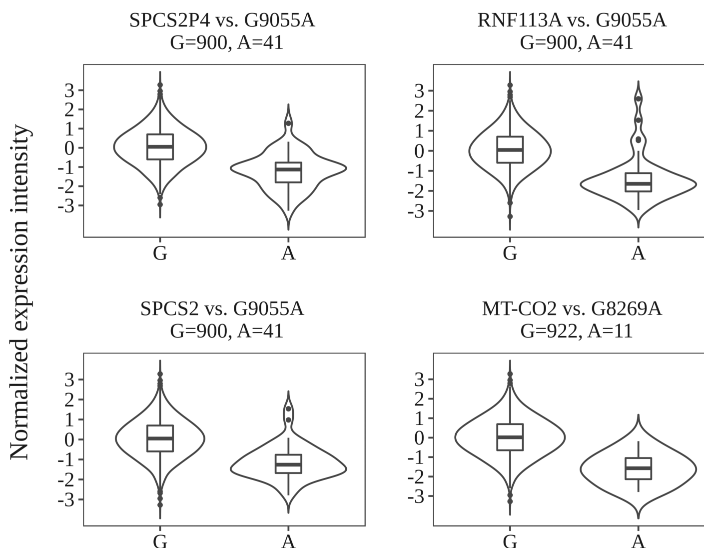

Figure 1. Combined boxplot and violin plot of the normalized expression intensities for the top four genes relative to the alleles of the top associated independent mtSNP.

compared to all other eight major haplogroups. The expression levels of the other three genes across haplogroups are shown in Supplementary Material, Figure S3.

The genes identified to be associated with mtSNPs or haplogroups and not located in the mitochondrial genome were *SPCS2P4*, *RNF113A*, *SPCS2*, *CRLS1*, *SLC25A14* and *SLC2A8*. None of these genes showed cross-hybridization with sequences on the mitochondrial genome.

Sex- and prediabetes-specific effects

A random-effect meta-analysis showed no statistically significant differences in the effect sizes of gene expression between the sexes (results not shown). The characteristics of the population used in the prediabetes-specific analysis are shown in Table 4. Age, sex and body mass index all differed significantly between the groups ($P \leq 0.001$). For one probe-mtSNP pair the meta-analysis showed a significant difference in the effect sizes between subjects with prediabetes and controls. A *P*-value of 8.91×10^{-9} ($\lambda_{GC} = 0.99$, Supplementary Material, Fig. S4) corresponded to the association between the expression of 2′-5′-oligoadenylate synthase like (*OASL*, Illumina Array Address ID 6280543) and mtSNP C16294T. Subjects with prediabetes had an effect estimate of −0.74 [standard error (SE) of 0.12] and a corresponding *P*-value of 9.69×10^{-6} ($\lambda_{GC} = 0.99$, Supplementary Material, Fig. S5), while the control group had an effect estimate of 0.43 (SE of 0.12) and a *P*-value of 4.22×10^{-4} ($\lambda_{GC} = 1.00$, Supplementary Material, Fig. S6). When we added an interaction term between C16294T and prediabetes status to the linear model explaining the expression of *OASL*, the interaction between the minor allele T and the presence of prediabetes was significant (effect estimate −1.07, SE 0.19 and a corresponding *P*-value of 1.11×10^{-8}). That is to say, on average, subjects with prediabetes and the minor allele T had lower expression levels of *OASL* compared to the reference allele, while subjects with the T allele but no prediabetes had higher expression levels compared to the reference allele, as can be seen in Figure 3. Basic Local Alignment Search Tool (BLAST) showed no evidence for cross-hybridization between *OASL* and sequences on the mitochondrial genome.

Downloaded from https://academic.oup.com/hmg/article/28/8/1381/5281396 by guest on 16 November 2020

Table 3. Haplogroup-wise comparisons from Tukey's post hoc test for the five differentially expressed genes identified in ANCOVA

Comparison	Difference in means [95% CI]	Tukey-adjusted P-value
SPCS2P4		
K-H	−1.27 [−1.80, −0.73]	1.33×10^{-11}
K-U	−1.23 [−1.78, −0.68]	2.38×10^{-10}
K-J	−1.30 [−1.93, −0.67]	8.57×10^{-9}
K-W	−1.35 [−2.04, −0.66]	6.08×10^{-8}
K-V	−1.17 [−1.82, −0.52]	9.93×10^{-7}
K-T	−1.10 [−1.78, −0.43]	1.55×10^{-5}
K-I	−1.47 [−2.38, −0.55]	2.74×10^{-5}
K-X	−1.28 [−2.14, −0.43]	1.09×10^{-4}
RNF113A		
K-H	−1.44 [−1.97, −0.91]	
K-U	−1.53 [−2.07, −0.99]	
K-J	−1.52 [−2.14, −0.90]	2.48×10^{-12}
K-T	−1.44 [−2.10, −0.77]	1.06×10^{-9}
K-W	−1.42 [−2.10, −0.74]	4.73×10^{-9}
K-X	−1.48 [−2.32, −0.65]	1.65×10^{-6}
K-I	−1.57 [−2.47, −0.67]	2.75×10^{-6}
K-V	−1.03 [−1.67, −0.39]	2.21×10^{-5}
K-U	−0.50 [−0.94, −0.06]	1.20×10^{-2}
SLC25A14		
W-J	0.81 [0.21, 1.41]	1.01×10^{-3}
W-H	0.55 [0.06, 1.05]	1.63×10^{-2}
K-J	0.69 [0.05, 1.33]	2.53×10^{-2}
W-X	0.86 [0.02, 1.69]	3.79×10^{-2}
SLC2A8		
U-J	−0.46 [−0.88, −0.04]	2.06×10^{-2}
U-H	−0.25 [−0.50, 0.00]	4.34×10^{-2}
MT-ND5		
J-V	0.84 [0.30, 1.39]	7.05×10^{-5}
J-U	0.59 [0.16, 1.01]	5.78×10^{-4}
J-H	0.50 [0.10, 0.91]	3.24×10^{-3}
J-W	0.69 [−1.29, −0.09]	1.14×10^{-2}

Abbreviation: CI, confidence interval
One unit of difference in means represents the proportion of one SD change in normalized gene expression intensity.

Table 4. Characteristics of the population used in the prediabetes-specific analysis. Values are means (SD)

	Controls	Individuals with prediabetes
Number of subjects in group	584	249
Age, years	41.6 (5.00)	42.8 (5.10)
Males (%)	202 (34.5)	143 (57.4)
Body mass index, kg/m^2	25.5 (4.16)	28.6 (5.42)

The Mann–Whitney U test was used for age and body mass index and the χ^2 test for sex when the two groups were compared. P-value $\leq$0.001 for all comparisons.

Discussion

In this study, we first wanted to replicate the earlier gene-mtSNP associations (7). We were able to replicate 6 of the 15 reported associations. The nucleotide sequence in these previous results differs from our sequence by one position, mtSNP A3481G in the previous sequence corresponds with A3480G in our sequence and A10551G with A10550G.

Secondly, we set out to find new associations with a greater number of mtSNPs by using population-based mtDNA sequencing. From 3 907 763 analyzed expression probe-mtSNP pairs, we identified 47 new associations after a strict correction for multiple testing. The majority of the associations included either pseudogene *SPCS2P4* in chromosome 1 or *RNF113A* in the X chromosome. The first is involved in the biosynthesis of the N-glycan precursor and transfer to a nascent protein, while the latter encodes a protein containing two zinc finger domains.

After applying pairwise conditional analysis to all mtSNPs showing at least modest linkage disequilibrium, five probe-mtSNP pairs showed evidence for independent signaling. In the expression of *SPCS2P4*, the effects of G9055A and A11251G/C15452T were in opposite directions which also support the independence of these signals. For *SLC25A14* and *CRLS1*, none of the signals remained independent. This could be explained by the high linkage disequilibrium (pairwise $r^2 > 0.89$

Downloaded from https://academic.oup.com/hmg/article/28/8/1381/5281396 by guest on 16 November 2020

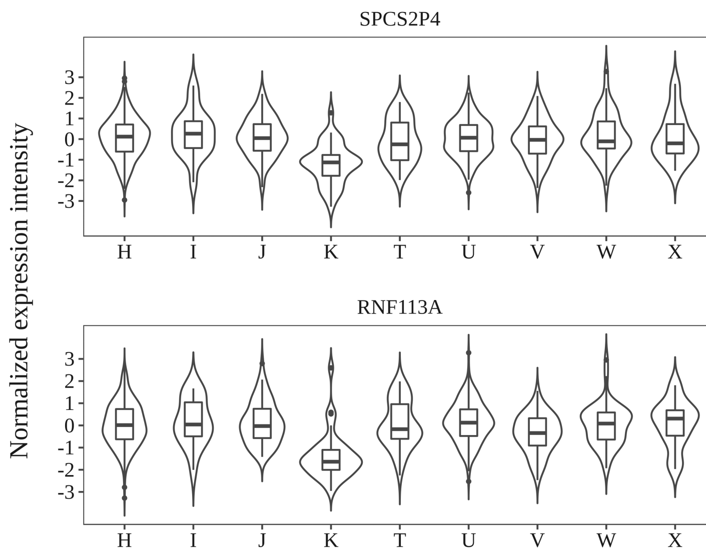

Figure 2. Combined boxplot and violin plot of the normalized expression intensities of *SPCS2P4* and *RNF113A* across the haplogroups.

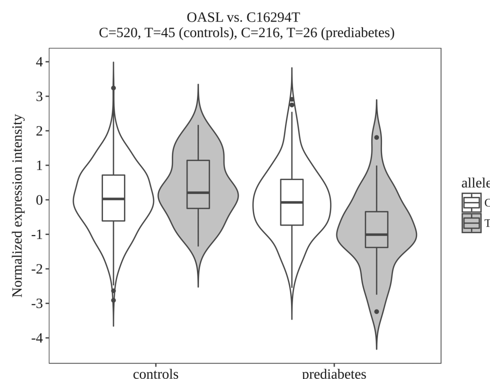

Figure 3. Combined boxplot and violin plot of the normalized expression intensities of *OASL* across the individuals with and without prediabetes relative to mtSNP C16294T.

Table 5. Absolute and relative mitochondrial haplogroup frequencies in the study population

Haplogroup	Absolute frequency	Relative frequency (%)
H	405	43.4
U	240	25.7
J	69	7.4
V	58	6.2
T	48	5.1
W	43	4.6
K	35	3.7
X	20	2.1
I	16	1.7

for all mtSNPs), all variants that associated with these two genes in the unconditioned analysis together predispose to differential gene expression. In addition, the associations with mitochondrial genes *MT-CO2* and *MT-RNR2L1* seem to result from independent signals. That is to say, in total, eight mtSNP-gene associations were independent.

We also discovered that the expression intensities of four nuclear and one mitochondrial gene differed between the major Scandinavian haplogroups present in our study population. The haplogroup distribution shown in Table 5 is similar to the one previously reported in Finnish population. Compared to other European populations, the frequency of haplogroup U appears to be higher and the frequency of haplogroup K lower in Finland (3,15). The lower intensities of *SPCS2P4* and *RNF113A* in haplogroup K result from the fact that the majority of the mtSNPs associated with these two genes (e.g. G9055A, A3480G and A10550G) are also the defining variants for this haplogroup (16). The expression intensities of two genes, *SLC2A8* and *MT-ND5*, were not associated at the mtSNP level. However, the majority of the mtSNPs (C295T, C462T, A12612G, G13708A, C16069T) defining haplogroup J (16) that associated with *MT-ND5*, were in the top 20 variants associating with this gene. Those associations were not however significant after correction for multiple testing. For haplogroup U that associated with SLC2A8, none of the defining variants were in top 20 mtSNPs associating with this gene. This implies that the haplogroup effect is driven by small individual effects of many mtSNPs, especially the expression of *SLC2A8*.

As already discussed earlier (7), the biological relevance of these reported associations remain unclear, and they do not necessarily imply causal relationships—i.e. all these mtSNPs are not necessarily expression regulatory SNPs. However, the associations may equally represent the altered cellular activities resulting from the mtSNPs that the nuclear genome

Downloaded from https://academic.oup.com/hmg/article/28/8/1381/5281396 by guest on 16 November 2020

is then compensating for (13). In other studies, some of the reported mtSNPs have been associated with non-mitochondrial diseases. For example, the variant G9055A and haplogroup K have been found to increase breast cancer risk in European–American women (17). Interestingly, increased levels of protein coded by *RNF113A* in plasma have been suggested to act as a biomarker for breast cancer (18), while our analysis showed a reduced expression of *RNF113A* associated with this mtSNP and haplogroup.

Our fourth aim was to investigate whether any probe-mtSNP associations show sex-specific differences by using a random-effect meta-analysis, which did not reveal any significant results. However, it is possible that sexually dimorphic effects on gene expression are mediated via other mtSNPs that were not included in this study. It is also worth mentioning that, to our knowledge, the only study that has previously examined these sex-specific effects in humans only took into account homoplasmic mtDNA alleles (7) which was also the case in the present study. For this reason, additional studies also considering heteroplasmic alleles are needed, although the link between heteroplasmy and sex-specific effects is not obvious.

Finally, we tested the hypothesis that probe-mtSNP associations are affected by the onset of prediabetes. We were able to find support for this hypothesis by comparing the effect sizes between samples with and without prediabetes, which showed that the onset of prediabetes affects the mitochondrial control of the expression of *OASL* through mtSNP C16294T. This mtSNP in the mtDNA control region has also been previously associated with cardiovascular risk factors, it has been linked to obesity in an Austrian population (19), and there is also evidence that C16294T is associated with coronary artery disease, although possibly through linkage to mtSNP T16189C (20). Functionally, *OASL* encodes an interferon-inducible antiviral protein, and its high expression levels in visceral adipose tissue, together with other three interferon signature genes, have been found to be positively correlated with adipose tissue and systemic insulin resistance (21). These previous results link *OASL* to prediabetes, since most persons with prediabetes are also insulin-resistant (22). Our results are in contrast to those published earlier in adipose tissue (21), since the expression of *OASL* was lower in T allele carriers with prediabetes compared those without prediabetes. However, profiling gene expression from peripheral blood leukocytes makes it challenging to speculate how the expression levels represent the expression in other tissues, such as adipose tissue. One hypothesis based on our results would be that the chronic low-grade inflammation in prediabetes (23) is, to some extent, milder in individuals with mtSNP C16294T. Whether this has any clinical significance remains to be examined in longitudinal studies. In addition, the effect of prediabetes on probe-mtSNP associations should be replicated in other data sets.

The prevalence of prediabetes in the non-diabetic Finnish population used in this study was 29.8%. In another Finnish cohort, the prevalence was 9.6% in men and 10.4% in women aged between 45 and 54 years (24). The higher prevalence in our study can partly be explained by the lower cut-off point for impaired fasting glucose. In addition, the study sample used in (24) included also those with T2D, whereas they were excluded in the current study. The current prevalence is in line with the one reported in the United States, where 35% of adults aged 20 years or older had prediabetes in 2005–2008 (25). Although the prevalence is high, 5–10% of individuals per year with prediabetes will progress to T2D, with the same proportion converting back to normoglycemia (26).

The strength of this study is that the mtSNPs were obtained through deep sequencing. Compared to microarray genotyping, this increased the number of mtSNPs to be included in the analyses. This study also has some limitations. The Finnish gene pool has been shown to be distinctive, and the results may not be directly generalizable to populations with a different ethnic background, a fact that is pronounced in mitochondrial genetic studies. Another limitation is that no oral glucose tolerance tests were performed on the study population and the definition of prediabetes was based only on fasting plasma glucose and HbA1c levels. However, the HbA1c cut-off point for prediabetes has a high specificity to identify cases of impaired glucose tolerance (14). We also recognize that microarray studies are limited by multiple testing problems and false positives, although the number of false positive results was minimized by using mitochondrial principal components (PCs) as covariates and a strict Bonferroni correction.

In summary, this study provides both novel and additional evidence for the mitochondrial genetic control of peripheral blood cell gene expression. No significant evidence of sex-specific effects of mtSNPs on gene expression was found, but we present evidence that the onset of prediabetes may lead to perturbations in this mitochondrial genetic control. The possible clinical relevance of these results remains to be examined in future functional and longitudinal studies.

Materials and Methods

Study participants

The Cardiovascular Risk in Young Finns Study (http://youngfinnsstudy.utu.fi) is a Finnish longitudinal population study on the evolution of cardiovascular risk factors from childhood to adulthood (27). We used data from the follow-up in 2011 when the subjects were aged between 34 and 49 years, with the exception of the mtDNA data, which was gathered in 2007. The follow-up studies in 2007 and 2011 included 2204 and 2060 participants, respectively. The study plan was approved by the ethics committees of all participating hospital districts, and the study protocol of each study phase corresponded with the proposal by the World Health Organization. All subjects gave written informed consent, and the study was conducted in accordance with the Declaration of Helsinki.

Blood transcriptomic analysis, RNA analysis and data processing

Whole-blood samples were obtained from 2049 individuals for RNA isolation. The 63 samples were discarded during the RNA isolation protocol, leading to 1987 samples including one technical replicate taking part in the genome-wide expression analysis. The 322 samples had too low concentration after amplification step. After this, 1667 samples (including three replicates, two from the mRNA amplification step) were analyzed with an Illumina HumanHT-12 version 4 Expression BeadChip (Illumina Inc.) containing 47 231 expression and 770 control probes. The transcripts detected (detection P-value < 0.01) in less than 5% of samples were excluded from the analysis. After this filtering 19 637 genes were used for analysis. We disregarded four samples with less than 6000 significantly detected expression probes (detection P-value < 0.01).

The expression data was processed in R (http://www.r-project.org/) using a nonparametric background correction, followed by quantile normalization with control and expression

Downloaded from https://academic.oup.com/hmg/article/28/8/1381/5281396 by guest on 16 November 2020

probes, using the neqc function in the limma package (28) and a log2 transformation. The expression levels were also zero centered, and rank-based inverse normal transformation was applied to further normalize the expression levels. Based on RPS4Y1–2 and XIST mRNA levels on the Y and X chromosomes, respectively, we excluded nine samples due to mismatch with the recorded sex. After quality control, expression data were available for 1654 samples including four technical replicates, which were used to examine batch effects and excluded subsequently. In summary, the expression analysis was successful for 19 637 probes and 1650 samples. Other details of the process have been described previously by Turpeinen *et al*. (29).

MtDNA sequencing

Genomic DNA sample (n = 1817) concentrations were measured from whole-blood samples with the Qubit BR dsDNA kit (Life Technologies Ltd). mtDNA was amplified from the genomic DNA using the REPLI-g mtDNA kit (Qiagen) in a 15 μl reaction volume. After the enrichment, the amplified mtDNA samples were processed into Illumina deep sequencing compatible libraries with the Nextera DNA sample preparation kit (Illumina Inc.). The mtDNA concentrations were measured with Qubit dsDNA for Nextera tagmentation reaction. The reaction volume in the Nextera tagmentation and amplification steps was 20 μl, and after both steps the libraries were purified with a EdgeBio Performa V3 96-Well Short Plate (Edge BioSystems). After the amplification, the libraries were first incubated with 4 μl of EdgeBio SOPE resin and then purified with EdgeBio Performa plates. After purification, 48 samples with different index tags were pooled together (2 μl each) in each pool and concentrated with DNA Clean & Concentrator^{TM}-5 (Zymo Research). The final volume of the concentrated pool was 15 μl. The sequencing-ready libraries were quantitated with an Agilent 2100 Bioanalyzer High Sensitivity kit (Agilent). The libraries were deep sequenced with the Illumina HiSeq system. All the samples (n = 1667) that achieved any mean bait coverage were included in the quality control process.

The primer sites in the REPLI-g kit have not been published and the data for each of the 16 samples was therefore validated by amplifying the mtDNA in two different Polymerase chain reaction (PCR) amplicons covering the whole mtDNA. The primers have been previously described by Pietiläinen *et al*. (30). The amplicons were processed into Illumina-compatible sequencing libraries according to the same Nextera protocol as detailed above. The data was analyzed using an in-house-developed bioinformatics pipeline (31). The variants from REPLI-g amplification and PCR amplification were compared, and no significant differences were observed. The comparison results confirm that REPLI-g primers do not affect the variant detection.

MtDNA quality control and data processing

First, samples with individual missingness >0.10 (134 samples) were excluded. Next, samples were classified into haplogroups by using HaploGrep (32) (Phylotree build 16) (16) after comparison to the revised Cambridge Reference Sequence (1). Only those individuals whose haplogroup quality score was above 0.90 were included for further analyses (66 samples rejected). At this quality threshold, haplogroup assignment is quite reliable, according to the HaploGrep manual. After this, 1467 samples remained for further analysis. Then, samples that had both gene expression and mtDNA data were merged, after which 955 samples were remaining. Next, 29 mtSNPs that obtained mean call ratios of <0.85 were excluded, all samples had mean call ratio above this value. After this, heteroplasmic alleles' status was set to missing. The remaining quality control steps included filtering for missingness by mtSNP >0.05 (34 mtSNPs discarded) and minor allele frequency (MAF) < 0.01 (538 mtSNPs discarded). After these procedures, a total of 199 mtSNPs from 955 samples (548 women and 407 men) were available for probe-mtSNP association analysis (Supplementary Material, Table S2).

For association testing, the haplogroups were assigned to major haplogroups. Haplogroups with a frequency of less than 0.01 were excluded, leaving 934 samples for the haplogroup-probe analysis. The haplogroup frequencies are shown in Table 5. All major Scandinavian haplogroups except haplogroup M were present (3).

Definition of prediabetes

Venous blood samples were drawn after an overnight fast for the determination of serum glucose and glycated hemoglobin A1c (HbA1c). The classification of prediabetes was based on fasting glucose and HbA1c according to the criteria of the American Diabetes Association (14). People with impaired fasting glucose were defined as having a fasting plasma glucose level of 5.6–6.9 mmol/L or an HbA1c level of 39–47 mmol/mol without a diagnosis of T2D. The diagnosis of T2D included subjects with a fasting plasma glucose level of 7.0 mmol/L or higher or an HbA1c level of 48 mmol/mol or higher or those who reported using oral glucose-lowering medication or insulin (but had not reported having type 1 diabetes) or who had a reported diagnosis of T2D by a physician. Those diagnosed with type 1 diabetes were also ruled out. Of the 833 subjects for whom gene expression and mtDNA data and prediabetes status were available, 249 had prediabetes, and 584 controls had normal levels of fasting plasma glucose and HbA1c.

Statistical analysis

In order to investigate the association of peripheral blood gene expression with mtSNPs, the expression levels were modeled as a linear function of the presence (coded as 1) or absence (coded as 0) of the minor allele using the lm function in R. Age and sex were added as covariates to the linear model. We calculated the P-values using a standard F test with one degree of freedom and accounted for multiple testing using the Bonferroni correction. Significance was defined as $P < 1.28 \times 10^{-8}$ (i.e. 0.05/[199 × 19 637]). With this P-value, MAF of 0.01 and a minimum statistical power of 0.80, the minimum detectable effect size was 1.58. With MAF of 0.05, the minimum effect size was 0.72 (33). ANCOVA with age and sex as covariates was employed to flag genes for those showing differential expression between haplogroups. All genes with a P-value of $<2.55 \times 10^{-6}$ (i.e. 0.05/19 637) were compared using Tukey's honest significant difference test to confirm the between-haplogroup differences. A Tukey-adjusted P-value of <0.05 was considered statistically significant.

The sex-specific effects of mtSNPs on gene expression were tested by applying the same linear model as described above (sex was removed from the covariates) to males and females separately. Differences in effect sizes were compared by applying a random-effect meta-analytic model to each probe using the MetaDE package (34). Heterogeneity was examined by Cochran's

Downloaded from https://academic.oup.com/hmg/article/28/8/1381/5281396 by guest on 16 November 2020

Q test with the corresponding P-value. A significant P-value would suggest that there is a significant difference in effect sizes between the sexes. For the sex-specific meta-analysis, the number of mtSNPs tested was 156 because for some mtSNPs the minor allele frequency was less than 0.01 in either males or females (Supplementary Material, Table S3). Significance was then defined as $P < 1.63 \times 10^{-8}$ (i.e. 0.05/[156 × 19 637]).

The effect of prediabetes on the association between mtSNPs and gene expression was studied similarly. Age, sex and body mass index were added as covariates to the linear regression models. The number of mtSNPs included was now 127, resulting in significance in random-effect meta-analysis defined as $P < 2.00 \times 10^{-8}$ (i.e. 0.05/[127 × 19 637]) (Supplementary Material, Table S4). The Mann–Whitney U test was used for age and body mass index and the χ^2 test for sex when the two groups were compared. After identifying the significant associations, we also studied the interaction between significant mtSNPs and prediabetes status on the expression of pinpointed genes by adding an interaction term to the linear model described above.

PC analysis was performed on all nuclear probes passing quality control. The prcomp function (package stats) was used to calculate nuclear PCs 1–20 from mean expression levels. The use of gene expression principal components should reduce the effect of technical factors (Illumina chip assignment, the RNA amplification batch, the RNA isolation batch and the sample storage time, in particular the time between blood donation and RNA isolation) in our data (35). The use of mitochondrial PCs has been demonstrated to be a robust method to control of confounding due to population stratification. In addition, the use of mitochondrial PCs effectively removes false positive associations but does not cause loss in power for detection of true associations (36). PC analysis was performed on all mtSNPs passing quality control. Since the mtSNP were represented as binary variables, i.e. the presence or absence of minor allele, logisticPCA package (37) was used to extract mitochon drial PCs 1–20.

Both nuclear and mitochondrial PCs were added as covariates, in addition to those mentioned above, in the linear mtSNP association models until no additional reduction in genomic inflation factor (λ_{GC}) could be achieved. Nuclear PCs 1–11 and mitochondrial PCs 1–2 were used for all probe-mtSNP association analyses. For haplogroup analysis in ANCOVA, nuclear PCs 1–11 were used. Mitochondrial PCs cannot be applied to haplogroup analysis, since the haplogroups are strongly correlated with the mitochondrial PCs. The genomic inflation factor was calculated using the GenABEL package (38). Values of $\lambda_{GC} < 1.05$ are generally considered benign (39).

For all significant mtSNPs for each probe, pairwise linkage disequilibrium between mtSNPs was quantified as squared Pearson correlation r^2. All significant mtSNPs in at least modest pairwise linkage disequilibrium ($r^2 > 0.30$) were subjected to pairwise conditional analysis in order to identify the independent signals. Conditional analysis was performed by using the same linear model as in the full analysis but additionally conditioning for one additional significant mtSNP at a time. We applied an mtSNP-specific Bonferroni correction (i.e. correction for the number of pairwise analyses made for each mtSNP) to account for multiple testing.

All significant probes from the above analyses were tested for cross-hybridization with sequences other than the target transcript using algorithm blastn from BLAST (40). We were particularly interested in probes that have sequence similarities with the mtDNA. Probes were considered to show strong evidence for cross-hybridization with the mtDNA if probes' sequences had 90% identity over the aligned region, at least 40 of 50 matching bps, and no gaps.

Data availability

Due to legal restrictions, the Ethics Committee of the Hospital District of Southwest Finland has stated that individual level data cannot be stored in public repositories or otherwise made publicly available. Data are, however, available from the authors upon a reasonable request.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We thank Jari Kaikkonen and Janne Asikainen for their technical assistance.

Conflict of Interest statement. The authors declare that there is no conflict of interest associated with this manuscript.

Funding

Juhani Aho Foundation and the Aarne Koskelo Foundation [to J.L.]; the Young Finns Study has been financially supported by the Academy of Finland [grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi) and 41071 (Skidi)]; the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals [grant X51001]; the Juho Vainio Foundation; the Paavo Nurmi Foundation; the Finnish Foundation for Cardiovascular Research; the Finnish Cultural Foundation; the Tampere Tuberculosis Foundation; the Emil Aaltonen Foundation; the Yrjö Jahnsson Foundation; the Signe and Ane Gyllenberg Foundation; the Diabetes Research Foundation of the Finnish Diabetes Association; the Finnish Cultural Foundation—The Pirkanmaa Regional Fund; European Union Horizon 2020 [grant 755320 for TAXINOMISIS].

References

1. 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. *Nat. Genet.*, **23**, 147.
2. Kivisild, T., Shen, P., Wall, D.P., Do, B., Sung, R., Davis, K.K., Passarino, G., Underhill, P.A., Scharfe, C., Torroni, A. *et al.* (2005) The role of selection in the evolution of human mitochondrial genomes. *Genetics*, **172**, 373–387.
3. 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**, 1835–1850.
4. Hurst, L.D. and Frank, S.A. (1996) Mitochondria and male disease. *Nature*, **383**, 224.
5. 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 Ecol. Evol.*, **19**, 238–244.
6. Innocenti, P., Morrow, E.H. and Dowling, D.K. (2011) Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. *Science*, **332**, 845–848.

Downloaded from https://academic.oup.com/hmg/article/28/8/1381/5281396 by guest on 16 November 2020

7. Kassam, I., Qi, T., Lloyd-Jones, L., Holloway, A., Bonder, M.J., Henders, A.K., Martin, N.G., Powell, J.E., Franke, L., Montgomery, G.W. *et al.* (2016) Evidence for mitochondrial genetic control of autosomal gene expression. *Hum. Mol. Genet.*, **25**, 5332–5338.
8. Woodson, J.D. and Chory, J. (2008) Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.*, **9**, 383–395.
9. Bar-Yaacov, D., Blumberg, A. and Mishmar, D. (2012) Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation. *Biochim. Biophys. Acta*, **1819**, 1107–1111.
10. Chinnery, P.F., Elliott, H.R., Patel, S., Lambert, C., Keers, S.M., Durham, S.E., McCarthy, M.I., Hitman, G.A., Hattersley, A.T. and Walker, M. (2005) Role of the mitochondrial DNA 16184-16193 poly-C tract in type 2 diabetes. *Lancet*, **366**, 1650–1651.
11. Park, K.S., Chan, J.C., Chuang, L.M., Suzuki, S., Araki, E., Nanjo, K., Ji, L., Ng, M., Nishi, M., Furuta, H. *et al.* (2008) A mitochondrial DNA variant at position 16189 is associated with type 2 diabetes mellitus in Asians. *Diabetologia*, **51**, 602–608.
12. Saxena, R., De Bakker, P.I.W., Singer, K., Mootha, V., Burtt, N., Hirschhorn, J.N., Gaudet, D., Isomaa, B., Daly, M.J., Groop, L. *et al.* (2006) Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am. J. Hum. Genet.*, **79**, 54–61.
13. Hwang, S., Kwak, S.H., Bhak, J., Kang, H.S., Lee, Y.R., Koo, B.K., Park, K.S., Lee, H.K. and Cho, Y.M. (2011) Gene expression pattern in transmitochondrial cytoplasmic hybrid cells harboring type 2 diabetes-associated mitochondrial DNA haplogroups. *PLoS One*, **6**, e22116.
14. American Diabetes Association (2010) Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, **33**, S62–S69.
15. Richards, M., Côrte-Real, H.C., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H.-J. and Sykes, B. (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.*, **59**, 185–203.
16. Van Oven, M. and Kayser, M. (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.*, **30**, E386–E394.
17. Bai, R.-K., Leal, S.M., Covarrubias, D., Liu, A. and Wong, L.-J.C. (2007) Mitochondrial genetic background modifies breast cancer risk. *Cancer Res.*, **67**, 4687–4694.
18. Li, C.I., Mirus, J.E., Zhang, Y., Ramirez, A.B., Ladd, J.J., Prentice, R.L., McIntosh, M.W., Hanash, S.M. and Lampe, P.D. (2012) Discovery and preliminary confirmation of novel early detection biomarkers for triple-negative breast cancer using preclinical plasma samples from the Women's Health Initiative observational study. *Breast Cancer Res. Treat.*, **135**, 611–618.
19. Ebner, S., Mangge, H., Langhof, H., Halle, M., Siegrist, M., Aigner, E., Paulmichl, K., Paulweber, B., Datz, C., Sperl, W. *et al.* (2015) Mitochondrial haplogroup T is associated with obesity in Austrian juveniles and adults. *PLoS One*, **10**, e0135622.
20. Mueller, E.E., Eder, W., Ebner, S., Schwaiger, E., Santic, D., Kreindl, T., Stanger, O., Paulweber, B., Iglseder, B., Oberkofler, H. *et al.* (2011) The mitochondrial T16189C polymorphism is associated with coronary artery disease in Middle European populations. *PLoS One*, **6**, e16455.
21. Ghosh, A.R., Bhattacharya, R., Bhattacharya, S., Nargis, T., Rahaman, O., Duttagupta, P., Raychaudhuri, D., Shiu, C., Liu, C., Roy, S. *et al.* (2016) Adipose recruitment and activation of plasmacytoid dendritic cells fuel metaflammation. *Diabetes*, **65**, 3440–3452.
22. Ferrannini, E., Gastaldelli, A. and Iozzo, P. (2011) Pathophysiology of prediabetes. *Med. Clin. North Am.*, **95**, 327–339.
23. Festa, A., Hanley, A.J.G., Tracy, R.P., D'agostino, R. and Haffner, S.M. (2003) Inflammation in the prediabetic state is related to increased insulin resistance rather than decreased insulin secretion. *Circulation*, **108**, 1822–1830.
24. Peltonen, M., Korpi-Hyövälti, E., Oksa, H., Puolijoki, H., Saltevo, J., Vanhala, M., Saaristo, T., Saarikoski, L., Sundvall, J. and Tuomilehto, J. (2006) Prevalence of obesity, type 2 diabetes, and other disturbances in glucose metabolism in Finland - FIN-D2D survey. *Finnish Med. J.*, **61**, 163–170.
25. Centers for Disease Control and Prevention (2011) National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. US Department of Health and Human Services, Centers for Disease Control and Prevention, Atlanta, GA,
26. Tabák, A.G., Herder, C., Rathmann, W., Brunner, E.J. and Kivimäki, M. (2012) Prediabetes: a high-risk state for diabetes development. *Lancet*, **379**, 2279–2290.
27. Juonala, M., Viikari, J.S. and Raitakari, O.T. (2013) Main findings from the prospective cardiovascular risk in young Finns study. *Curr. Opin. Lipidol.*, **24**, 57–64.
28. Smyth, G.K. (2005) Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, New York, NY, pp. 397–420.
29. Turpeinen, H., Seppälä, I., Lyytikäinen, L.L.-P., Raitoharju, E., Hutri-Kähönen, N., Levula, M., Oksala, N., Waldenberger, M., Klopp, N., Illig, T. *et al.* (2015) A genome-wide expression quantitative trait loci analysis of proprotein convertase subtilisin/kexin enzymes identifies a novel regulatory gene variant for FURIN expression and blood pressure. *Hum. Genet.*, **134**, 627–636.
30. Pietiläinen, K.H., Naukkarinen, J., Rissanen, A., Saharinen, J., Ellonen, P., Kernen, H., Suomalainen, A., Götz, A., Suortti, T., Yki-Järvinen, H. *et al.* (2008) Global transcript profiles of fat in monozygotic twins discordant for BMI: pathways behind acquired obesity. *PLoS Med.*, **5**, e51.
31. Sulonen, A.-M., Ellonen, P., Almusa, H., Lepistö, M., Eldfors, S., Hannula, S., Miettinen, T., Tyynismaa, H., Salo, P., Heckman, C. *et al.* (2011) Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol.*, **12**, R94.
32. 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. *Hum. Mutat.*, **32**, 25–32.
33. Dong, X., Chang, T.-W., Weiss, S.T. and Qiu, W. (2017) powerEQTL: power and sample size calculation for eQTL analysis. R package version 0.1.3.
34. Wang, X., Kang, D.D., Shen, K., Song, C., Lu, S., Chang, L.-C., Liao, S.G., Huo, Z., Tang, S., Ding, Y. *et al.* (2012) An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. *Bioinformatics*, **28**, 2534–2536.
35. Schurmann, C., Heim, K., Schillert, A., Blankenberg, S., Carstensen, M., Dörr, M., Felix, S., Endlich, K., Gieger, C., Grallert, H. *et al.* (2012) Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaXpress consortium. *PLoS One*, **7**, e50938.
36. Biffi, A., Anderson, C.D., Nalls, M.A., Rahman, R., Sonni, A., Cortellini, L., Rost, N.S., Matarin, M., Hernandez, D.G., Plourde, A. *et al.* (2010) Principal-component analysis for assessment

Downloaded from https://academic.oup.com/hmg/article/28/8/1381/5281396 by guest on 16 November 2020

of population stratification in mitochondrial medical genetics. *Am. J. Hum. Genet.*, **86**, 904–917.

37. Landgraf, A.J. and Lee, Y. (2015) Dimensionality reduction for binary data through the projection of natural parameters. Technical Report No. 890, Department of Statistics, The Ohio State University, Columbus, OH.
38. Aulchenko, Y.S., Ripke, S., Isaacs, A. and Van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, **23**, 1294–1296.
39. Price, A.L., Zaitlen, N.A., Reich, D. and Patterson, N. (2010) New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.*, **11**, 459–463.
40. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Downloaded from https://academic.oup.com/hmg/article/28/8/1381/5281396 by guest on 16 November 2020

PUBLICATION
II

Mitochondrial genome-wide analysis of nuclear DNA methylation quantitative trait loci

Laaksonen, J., Mishra, P. P., Seppälä, I., Raitoharju, E., Marttila, S., Mononen, N., Lyytikäinen, L.-P., Kleber, M. E., Delgado, G. E., Lepistö, M., Almusa, H., Ellonen, P., Lorkowski, S., März, W., Hutri-Kähönen, N., Raitakari, O., Kähönen, M., Salonen, J. T. and Lehtimäki, T.

Human Molecular Genetics 31.10 (2022), 1720–1732

Publication reprinted with the permission of the copyright holders

OXFORD

Human Molecular Genetics, 2022, Vol. 31, 10, 1720–1732

https://doi.org/10.1093/hmg/ddab339
Advance access publication date 19 November 2021
Association Studies Article

Mitochondrial genome-wide analysis of nuclear DNA methylation quantitative trait loci

Jaakko Laaksonen[1,*], Pashupati P. Mishra[1], Ilkka Seppälä[1], Emma Raitoharju[1,2], Saara Marttila[2,3], Nina Mononen[1], Leo-Pekka Lyytikäinen[1], Marcus E. Kleber[4], Graciela E. Delgado[4], Maija Lepistö[5], Henrikki Almusa[5], Pekka Ellonen[5], Stefan Lorkowski[6,7], Winfried März[4,7,8,9], Nina Hutri-Kähönen[10], Olli Raitakari[11,12,13], Mika Kähönen[14,15], Jukka T. Salonen[16,17] and Terho Lehtimäki[1]

[1]Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33520, Finland
[2]Molecular Epidemiology, Faculty of Medicine and Health Technology, Tampere University, Tampere 33520, Finland
[3]Gerontology Research Center, Tampere University, Tampere 33520, Finland
[4]Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim 68167, Germany
[5]Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki 00290, Finland
[6]Institute of Nutritional Sciences, Friedrich Schiller University Jena, Jena 07743, Germany
[7]Competence Cluster for Nutrition and Cardiovascular Health (nutriCARD) Halle-Jena-Leipzig, Jena 07743, Germany
[8]SYNLAB Academy, SYNLAB Holding Deutschland GmbH, Augsburg 86156, Germany
[9]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz 8010, Austria
[10]Tampere Centre for Skills Training and Simulation, Tampere University, Tampere 33520, Finland
[11]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku 20520, Finland
[12]Research Centre for Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland
[13]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20520, Finland
[14]Department of Clinical Physiology, Tampere University Hospital, Tampere 33520, Finland
[15]Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33520, Finland
[16]Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki 00014, Finland
[17]MAS-Metabolic Analytical Services Oy, Helsinki 00990, Finland

*To whom correspondence should be addressed at: Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Arvo Ylpön katu 34, PO Box 100, Tampere FI-33014, Finland. Tel: +358 504080774; E-mail: jaakko.h.laaksonen@tuni.fi

Abstract

Mitochondria have a complex communication network with the surrounding cell and can alter nuclear DNA methylation (DNAm). Variation in the mitochondrial DNA (mtDNA) has also been linked to differential DNAm. Genome-wide association studies have identified numerous DNAm quantitative trait loci, but these studies have not examined the mitochondrial genome. Herein, we quantified nuclear DNAm from blood and conducted a mitochondrial genome-wide association study of DNAm, with an additional emphasis on sex- and prediabetes-specific heterogeneity. We used the Young Finns Study ($n = 926$) with sequenced mtDNA genotypes as a discovery sample and sought replication in the Ludwigshafen Risk and Cardiovascular Health study ($n = 2317$). We identified numerous significant associations in the discovery phase ($P < 10^{-9}$), but they were not replicated when accounting for multiple testing. In total, 27 associations were nominally replicated with a $P < 0.05$. The replication analysis presented no evidence of sex- or prediabetes-specific heterogeneity. The 27 associations were included in a joint meta-analysis of the two cohorts, and 19 DNAm sites associated with mtDNA variants, while four other sites showed haplogroup associations. An expression quantitative trait methylation analysis was performed for the identified DNAm sites, pinpointing two statistically significant associations. This study provides evidence of a mitochondrial genetic control of nuclear DNAm with little evidence found for sex- and prediabetes-specific effects. The lack of a comparable mtDNA data set for replication is a limitation in our study and further studies are needed to validate our results.

Introduction

Mitochondrial DNA (mtDNA) encodes 22 transfer RNAs, two ribosomal RNAs and 13 protein subunits of the 4 oxidative phosphorylation (OXPHOS) complexes (1). The mutation rate of mtDNA is significantly higher than that of nuclear DNA, and mitochondrial single-nucleotide polymorphisms (mtSNPs) have accumulated during evolution, dividing the human population into mitochondrial haplogroups just as populations have colonized different geographic areas of the world (2).

Most of the mitochondrial proteome is encoded by nuclear DNA, and crosstalk between mitochondria and the nucleus is essential to maintaining normal cellular function. Retrograde signals from mitochondria to the nucleus induce changes in, for example, nuclear DNA methylation (DNAm) and gene expression, which, in

Received: May 22, 2021. **Revised:** November 12, 2021. **Accepted:** November 16, 2021
© The Author(s) 2021. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

turn, can regulate mitochondrial functionality and metabolism (3,4). Previous cohort-level studies have shown that mtSNPs and haplogroups also associate with nuclear gene expression in peripheral blood (5,6). If the associations arose from causal relationships, they could have been mediated by epigenetic changes. This hypothesis is backed up by an *in vitro* study carried out on human retinal cell cybrids with identical nuclei but different mtDNA, which demonstrated expression differences in inflammation, angiogenesis and signaling genes between different haplogroups (7). After treatment with a methylation inhibitor, the expression levels of these genes became equivalent. Also, alterations in the global DNAm levels have been identified between haplogroups in peripheral blood (8). However, the effect of individual mtSNPs on DNAm is less known, and cohort-level association studies are lacking. Genome-wide association studies have identified numerous DNAm quantitative trait loci (9–11), but these studies have not examined the mitochondrial genome.

Epigenetics may play a role in the etiology of type 2 diabetes (T2D) mellitus (12,13), and there is evidence that epigenetic changes are likely to be an early process that may occur before the onset of T2D, i.e. during prediabetes (14). Although mtSNPs and haplogroups do not seem to be associated with prediabetes or T2D in the European population (15–17), they may have smaller consequences on a molecular level or modulate the complications of the disease (18). For example, we have demonstrated that the onset of prediabetes may lead to changes in the mitochondrial genetic control of the peripheral blood transcriptome (6). However, the crosstalk between mtDNA and the nuclear epigenome in the setting of prediabetes is not known.

In the current study, we examined the mitochondrial genetic determinants of peripheral blood DNAm obtained from 926 participants in the Young Finns Study (YFS), with an additional focus on sex- and prediabetes-specific effects. We sought replication in an independent data set consisting of 2317 individuals from the Ludwigshafen Risk and Cardiovascular Health (LURIC) study and combined the replicated results in a meta-analysis. Finally, we studied the associations of the identified CpG sites with peripheral blood gene expression to explore possible biological consequences of the differential DNAm.

Results

Study characteristics

Table 1 provides the basic characteristics for both cohorts. The LURIC study participants were, on average, older than the YFS participants, with a higher percentage of men and individuals with prediabetes. The proportion of current smokers was similar in both cohorts, but the percentage of never-smokers was higher in the YFS. The fraction of ex-smokers in LURIC participants was also higher in every subgroup, except among women.

mtSNPs associated with DNAm

A total of 88 513 545 CpG–mtSNP pairs were tested in the discovery phase. The number of significant associations after accounting for multiple hypothesis testing ($P < 7.8 \times 10^{-10}$) was 5652, corresponding to 4618 unique CpG sites and 89 mtSNPs. The CpG sites were scattered all around the nuclear genome. A mitochondrial Manhattan plot representing the significant associations for all CpG sites is shown in Supplementary Material, Figure S1. The full list of significant associations is available as Supplementary Dataset S1. The bacon-adjusted values from the 88 513 545 CpG–mtSNP pairs yielded an estimated inflation factor (λ) of 1.00, which suggests minimal inflation.

In all, 685 CpG–mtSNP pairs that were significant in the discovery phase were available for replication, resulting in a significance level of $P < 7.3 \times 10^{-5}$ (0.05/685). None of the associations in the replication sample passed this threshold, even though we expected to see 228 associations to reach this *P*-value. Twenty-one associations were replicated with nominal significance ($P < 0.05$) (Table 2). At this threshold, we expected virtually all 685 associations to replicate. There was no correlation between the discovery and replication effect sizes (Pearson's $r = 0.06$, Supplementary Material, Fig. S2), and 51% of the associations had a consistent direction of effect. The fixed-effect meta-analysis combining the nominally replicated results yielded 19 associations with epigenome-wide significance ($P < 7.8 \times 10^{-10}$) (Table 2 and Fig. 1).

Sexual dimorphism

In the YFS, fixed-effect meta-analysis revealed significant differences in the effect sizes between the sexes for 664 CpG–mtSNP pairs, corresponding to 35 unique mtSNPs and 621 CpG sites, nine of which were located on the X chromosome (Supplementary Material, Dataset S2). Inflation of the results was minimal in both the male- and female-specific analyses (λ = 1.00 for both sexes). In the LURIC study, 135 of the 664 associations were available for replication. In all, 46% of the associations had a consistent direction of effect and there was no correlation of effect sizes between the discovery and replication cohorts (Pearson's $r = -0.08$, Supplementary Material, Fig. S3). None of the 135 associations exhibited sex-specific heterogeneity with $P < 3.7 \times 10^{-4}$ (0.05/135) or with $P < 0.05$.

Prediabetes-specific effects

In the discovery phase, 483 CpG–mtSNP pairs demonstrated a significant difference in the effect sizes between individuals with prediabetes and controls, corresponding to 470 unique CpGs and 26 mtSNPs (Supplementary Material, Dataset S3). No inflation was observed (λ = 1.00 for both groups). For replication, 113 CpG–mtSNP pairs were available, none of which were replicated with heterogeneity $P < 4.4 \times 10^{-4}$ (0.05/113) or with $P < 0.05$. No correlation of effect sizes between the two cohorts was observed (Pearson's $r = -0.04$,

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

Table 1. Basic characteristics of the YFS and LURIC cohorts; values are mean (SD) or *n* (%) for continuous and categorical variables, respectively

	All	Men	Women	Prediabetes	Controls
YFS					
No. of participants	926	401	525	263	597
Age, years	41.9 (5.1)	42.1 (5.1)	41.8 (5.1)	43.0 (5.1)	41.4 (5.0)
Women	525 (56.7)	–	–	104 (39.5)	385 (64.5)
BMI, kg/m^2	26.6 (5.0)	27.4 (4.6)	26.1 (5.1)	28.3 (5.4)	25.5 (4.2)
Active smoker	127 (13.7)	66 (16.5)	61 (11.6)	47 (17.9)	72 (12.1)
Smokes once a week or more often but not daily	34 (3.7)	19 (4.7)	15 (2.9)	8 (3.0)	22 (3.7)
Smokes less often than once a week	36 (3.9)	16 (4.0)	20 (3.8)	10 (3.8)	21 (3.5)
Attempts to quit smoking	12 (1.3)	7 (1.7)	5 (1.0)	4 (1.5)	7 (1.2)
Has quit smoking	234 (25.3)	111 (27.7)	123 (23.4)	71 (27.0)	150 (25.1)
Has never smoked	483 (52.2)	182 (45.4)	301 (57.3)	123 (46.8)	325 (54.5)
LURIC					
No. of participants	2317	1599	718	1105	311
Age, years	62.8 (10.7)	62.0 (10.6)	64.8 (10.5)	62.0 (10.8)	57.6 (12.4)
Women	718 (31.0)	–	–	328 (29.7)	98 (31.5)
BMI, kg/m^2	27.4 (4.1)	27.5 (3.8)	27.3 (4.7)	27.2 (3.8)	26.0 (3.8)
Heavy smokers	317 (13.7)	251 (15.7)	66 (9.2)	149 (13.5)	55 (17.7)
Light smokers	200 (8.6)	142 (8.9)	58 (8.1)	92 (8.3)	37 (11.9)
Former smokers, quit <10 years ago	319 (13.8)	254 (15.9)	65 (9.1)	154 (13.9)	43 (13.8)
Former smokers, quit ≥10 years ago	634 (27.4)	574 (35.9)	60 (8.4)	292 (26.4)	69 (22.2)
Has never smoked	847 (36.6)	378 (23.6)	469 (65.3)	418 (37.8)	107 (34.4)

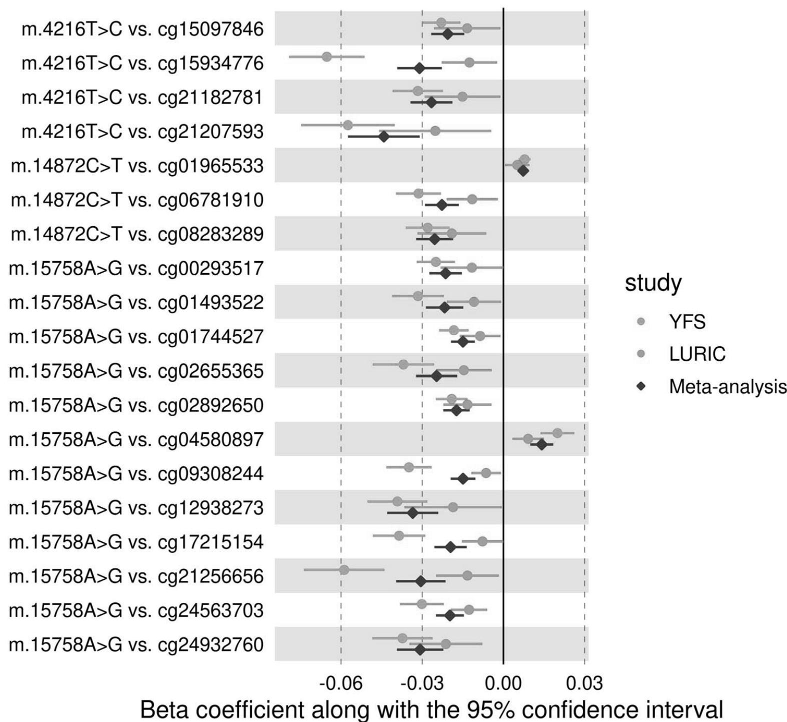

Figure 1. Forest plot showing the 19 nominally replicated mtSNP effects on DNAm, which also reached epigenome-wide significance in the meta-analysis.

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

Table 2. List of the 21 nominally replicated significant CpG–mtSNP associations; for all LURIC associations, the corresponding mtSNP was genotyped only with the HumanExome-12 array; 19 associations reached epigenome-wide significance in the fixed-effect meta-analysis ($P < 7.8 \times 10^{-10}$)

mtSNP	VAF	CpG	Chr	UCSC reference gene	YFS			LURIC			Meta-analysis		
					Effect	SE	P-value	Effect	SE	P-value	Effect	SE	P-value
m.499G>A	0.023	cg10790723	22	*FOXRED2*	0.030	0.005	5.1×10^{-10}	0.002	0.001	1.7×10^{-2}	0.002	0.001	1.1×10^{-3}
m.4216T>C[a]	0.12	cg15934776	2	*AFF3*	−0.065	0.007	4.4×10^{-20}	−0.013	0.005	1.6×10^{-2}	−0.031	0.004	1.8×10^{-13}
m.4216T>C[a]	0.12	cg21182781	1	*IARS2*	−0.032	0.005	2.6×10^{-11}	−0.015	0.007	3.5×10^{-2}	−0.027	0.004	1.8×10^{-11}
m.4216T>C[a]	0.12	cg21207593	17	*LIG3*	−0.057	0.009	7.5×10^{-11}	−0.025	0.011	1.7×10^{-2}	−0.044	0.007	7.1×10^{-11}
m.4216T>C[a]	0.12	cg15097846	12	*TULP3*	−0.023	0.004	1.8×10^{-10}	−0.013	0.006	3.3×10^{-2}	−0.021	0.003	4.4×10^{-11}
m.14872C>T[b]	0.020	cg06781910	4	–	−0.030	0.004	1.5×10^{-13}	−0.010	0.005	1.7×10^{-2}	−0.023	0.003	1.1×10^{-12}
m.14872C>T[b]	0.020	cg01965533	14	*DLST*	0.008	0.001	2.6×10^{-12}	0.005	0.002	2.7×10^{-2}	0.007	0.001	3.9×10^{-13}
m.14872C>T[b]	0.020	cg08283289	13	*ZC3H13*	−0.028	0.004	1.4×10^{-11}	−0.020	0.006	3.4×10^{-3}	−0.025	0.003	3.5×10^{-13}
m.15758A>G	0.014	cg09308244	2	–	−0.035	0.004	2.9×10^{-16}	−0.006	0.003	2.0×10^{-2}	−0.015	0.002	1.6×10^{-10}
m.15758A>G	0.014	cg17215154	7	*IQCE*	−0.039	0.005	5.7×10^{-15}	−0.008	0.004	4.7×10^{-2}	−0.020	0.003	1.7×10^{-10}
m.15758A>G	0.014	cg21256656	19	*KLK6*	−0.059	0.008	8.8×10^{-15}	−0.013	0.006	2.5×10^{-2}	−0.031	0.005	6.3×10^{-11}
m.15758A>G	0.014	cg24563703	4	–	−0.030	0.004	3.0×10^{-13}	−0.013	0.003	2.1×10^{-4}	−0.020	0.002	6.1×10^{-14}
m.15758A>G	0.014	cg00293517	3	–	−0.025	0.004	3.6×10^{-12}	−0.012	0.006	4.9×10^{-2}	−0.021	0.003	3.5×10^{-12}
m.15758A>G	0.014	cg08251499	8	*ZFAND1*	0.032	0.005	4.0×10^{-12}	0.002	0.001	5.8×10^{-3}	0.003	0.011	1.0×10^{-4}
m.15758A>G	0.014	cg12938273	22	*GNAZ; RSPH14*	−0.039	0.006	4.3×10^{-12}	−0.019	0.009	4.3×10^{-2}	−0.034	0.005	3.5×10^{-12}
m.15758A>G	0.014	cg01744527	19	*SCAF1*	−0.018	0.003	4.4×10^{-11}	−0.009	0.004	2.5×10^{-2}	−0.015	0.002	2.9×10^{-11}
m.15758A>G	0.014	cg24932760	17	*C17orf67*	−0.037	0.006	5.6×10^{-11}	−0.021	0.007	2.0×10^{-3}	−0.031	0.004	2.3×10^{-12}
m.15758A>G	0.014	cg01493522	13	–	−0.032	0.005	9.5×10^{-11}	−0.011	0.005	3.4×10^{-2}	−0.022	0.004	7.7×10^{-10}
m.15758A>G	0.014	cg02892650	16	*LRRC36*	−0.019	0.003	1.5×10^{-10}	−0.013	0.005	3.5×10^{-3}	−0.017	0.003	3.5×10^{-12}
m.15758A>G	0.014	cg02655365	6	*STK38*	−0.037	0.006	1.7×10^{-10}	−0.015	0.005	5.5×10^{-3}	−0.025	0.004	2.2×10^{-10}
m.15758A>G	0.014	cg04580897	3	*SNORA6; RPSA*	0.020	0.003	4.1×10^{-10}	0.009	0.003	2.1×10^{-3}	0.014	0.001	7.7×10^{-11}

Abbreviations and definitions: VAF, variant allele frequency in the discovery sample; Chr, chromosome; SE, standard error. [a]Tagged mtSNP m.15452C>A used in LURIC. [b]Tagged mtSNP m.2259C>T used in LURIC.

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

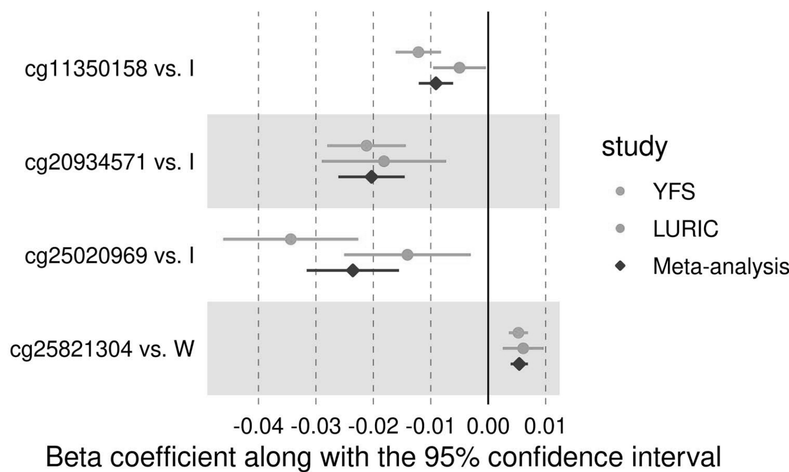

Figure 2. Forest plot showing the four nominally replicated haplogroup effects on DNAm, which also reached epigenome-wide significance in the meta-analysis. In all associations, haplogroup H was used as the reference haplogroup.

Supplementary Material, Fig. S4), and 50% of the associations had a consistent direction of effect.

Haplogroups associated with DNAm

The haplogroup frequencies and the corresponding phenotype characteristics of both cohorts are shown in Supplementary Material, Table S1. In both cohorts, the most common major haplogroup was H. In the discovery phase, a haplogroup-based analysis identified 142 significant associations (Supplementary Material, Dataset S4) with minimal inflation ($\lambda = 0.99$). The differentially methylated CpG sites were associated with six haplogroups: I (58.5% of the associations), X (22.5%), W (9.2%), K (4.9%), T (4.2%) and J (0.7%).

Twenty-two of the CpG sites that showed differential methylation in the YFS were not available in LURIC, leaving 120 CpG–haplogroup pairs for replication and setting the significance threshold at $P < 4.2 \times 10^{-4}$ (0.05/120). None of the associations in the LURIC survived this threshold; 15 associations were expected to reach this level. Six associations were nominally replicated, with the strongest association corresponding to $P = 8.2 \times 10^{-4}$. At a nominal threshold of $P < 0.05$, we expected all associations to replicate. There was a weak correlation between the discovery and replication effect sizes (Pearson's $r = 0.23$, Supplementary Material, Fig. S5), with 62% of the associations showing a consistent direction of effect. Four associations were significant in the meta-analysis ($P < 1.0 \times 10^{-8}$) (Table 3 and Fig. 2).

Expression quantitative trait methylation analysis

Overall, the replication phase identified 27 CpGs that showed differential methylation between mtSNPs or haplogroups. We considered genes ±1 Mb from each CpG site and tested 890 gene–CpG combinations for differential expression. Two associations were significant after correction for multiple testing: inverse associations were observed for cg25020969 (which showed lower methylation levels in haplogroup I) and probes ILMN_1681674 and ILMN_2358069, both at the *MAD1L1* gene (effect estimate: −4.63 and − 3.68, standard error: 1.05 and 0.91, P-value: 1.2×10^{-5} and 5.9×10^{-5}, respectively).

Discussion

The aim of the current study was to examine whether mtDNA variants and haplogroups associate with peripheral blood DNAm. Although previous studies have investigated the effect of haplogroups in cybrid cell lines or by using smaller sample sizes, the present study is, to the best of our knowledge, the first to examine the associations of mtSNPs on a cohort level and the largest to investigate the haplogroups' effects. In the discovery analysis of a Finnish population-based cohort, we identified numerous significant associations suggesting mitochondrial genetic control of DNAm and even pinpointed associations showing sex- and prediabetes-specific heterogeneity. Twenty-seven associations were nominally replicated in a German hospital-based cohort. The nominally replicated results were included in a joint meta-analysis of the two cohorts after which 23 associations remained significant. We were able to attempt replication only for approximately 15 – 30% of the mtSNP-based associations from the discovery phase, mainly owing to different mtDNA genotyping methods. We observed significantly lower replication rates than predicted even when we used a relatively liberal replication threshold ($P < 0.05$). The lack of replication was not explained by the winner's curse. Study-specific heterogeneity owing to different cohort characteristics may have had a major impact on

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

Table 3. List of nominally replicated CpG-haplogroup associations; four associations were significant in the meta-analysis ($P < 1.0 \times 10^{-8}$)

Haplogroup	CpG	Chr	UCSC reference gene	YFS			LURIC			Meta-analysis		
				Effect	SE	P-value	Effect	SE	P-value	Effect	SE	P-value
W	cg25821304	17	*RNF135*	0.005	0.001	6.1×10^{-10}	0.006	0.001	8.2×10^{-4}	0.005	0.001	2.2×10^{-12}
I	cg20934571	13	*CARKD*	−0.021	0.003	1.3×10^{-9}	−0.018	0.006	1.0×10^{-3}	−0.020	0.003	5.9×10^{-12}
I	cg11350158	2	*LRP1B*	−0.012	0.002	1.7×10^{-9}	−0.005	0.002	3.3×10^{-2}	−0.009	0.001	2.6×10^{-9}
I	cg24280540	8	–	0.030	0.005	4.6×10^{-9}	0.009	0.004	1.8×10^{-2}	0.017	0.003	5.7×10^{-8}
I	cg04933492	16	–	−0.028	0.005	5.8×10^{-9}	−0.008	0.004	4.4×10^{-2}	−0.015	0.002	2.7×10^{-7}
I	cg25020969	7	*MAD1L1*	−0.034	0.006	9.8×10^{-9}	−0.014	0.006	1.3×10^{-2}	−0.024	0.004	9.3×10^{-9}

replication. In addition, most of the effects identified in the YFS population may simply not be present among the LURIC study participants.

The most promising association was the one between haplogroup W and CpG site cg25821304, mapping to the gene *RNF135* (Table 3). It was significant in the discovery phase, reached borderline significance in replication and was significant in the meta-analysis. This CpG site did not show significant mRNA transcript associations, complicating the interpretation of the functional relevance. It has been documented that mitochondria are important participants in innate immune responses to pathogens and cellular damage and that also mtDNA variation could influence those immune response pathways (19). Our finding provides suggestive evidence for this since the protein encoded by the *RNF135* gene is involved in the evoking of innate immunity against RNA virus infections (20).

Of all the 27 identified CpG sites, 2 were significant in the expression quantitative trait methylation analysis. These associations, both corresponding to cg25020969, were not surprising, as the target gene of this CpG site and the two associated transcripts were the same, *MAD1L1*. It should be noted, however, that rather than being a strict dynamic mechanism for regulating gene expression, DNAm changes can also serve as a long-term memory of previous gene expression decisions that were mediated by transcriptional factors that might no longer be present in the cell (21).

The retrograde signals through which mitochondria affect nuclear DNAm appear to be mediated by tricarboxylic acid (TCA) cycle metabolites (22,23). α-Ketoglutarate (α-KG) serves as a cofactor for ten-eleven translocation hydroxylases (TET1-3) involved in DNA demethylation, whereas fumarate and succinate inhibit the TET enzymes. Even though the enzymes involved in the TCA cycle are not encoded by mtDNA, the TCA cycle is in constant feedback with the OXPHOS complexes, providing a plausible link between mtDNA variation, TCA metabolites and DNAm (22). In addition, experimental findings have directly coupled mtDNA variation with TCA metabolites and histone methylation (24). The three DNA methyltransferases (DNMT1, DNMT3A and DNMT3B) use S-adenosyl methionine (SAM) as a methyl donor. Although SAM is generated by coupling the methionine and folate cycles in the cytosol, these cycles are dependent on intermediate mitochondrial metabolism and ATP, and therefore, mtDNA variation may affect the function of DNMTs (3). To support these hypotheses, experimental findings on mouse embryonic stem cells have shown that mitochondrial haplogroups modulate the key regulators of both DNAm and demethylation, DNMT1 and TET1, leading to haplogroup-specific DNAm and gene expression patterns (25).

The association between haplogroup I and cg20934571 (Table 3) may represent the retrograde response aiming to regulate mitochondrial function. The CpG annotates to

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

NADPHX dehydratase (*NAXD*, also known as *CARKD*), and the protein product may be targeted to the mitochondria. The protein catalyzes the repair of NADPHX, a damaged form of reduced nicotinamide adenine dinucleotide phosphate (NADPH) (26), and mitochondrial NADPH plays a critical role in protecting the cells against mitochondrial oxidative stress (27). Based on this, it could be suggested that mutations defining haplogroup I result in disturbances in NADPH homeostasis, which leads to compensatory epigenetic changes. Another interesting association is between variant m.14872C>T (in the *MT-CYB* gene, a subunit of OXPHOS complex III) and cg01965533 (Table 2), which annotates to dihydrolipoamide succinyltransferase (*DLST*). The protein product of *DLST* is a subunit of the α-KG dehydrogenase complex, which is a key control point in the TCA cycle (28). Even though complex III is not directly coupled to the TCA cycle, the identified association could result from mitochondrial–nuclear communication, such as alterations in the electron transport chain that are compensated by epigenetic changes. However, these speculations are purely hypothetical and these two CpGs sites did not associate with mRNA transcripts.

We did not observe differential methylation at CpG sites mapping to the genes that showed mtSNP- or haplogroup-specific transcriptome profiles in Kassam *et al.* (5) or in our previous study (6). In addition, variant m.3480A>G was strongly associated with nuclear DNA transcripts in both of these previous studies but was not significantly associated with DNAm at any stage in the present study. This suggests that, if there is a causal relationship between mtSNP m.3480A>G (and the mtSNPs tagged by it) and peripheral blood transcriptomics, the expression regulatory mechanisms are not mediated by changes in DNAm.

In a study using articular cartilage cells (29), haplogroup J was associated with differentially methylated CpG sites when compared with haplogroup H. We could not validate these results, as only one CpG site was differentially associated with haplogroup J in the discovery phase and none during replication. The present and the aforementioned study had different sample sizes and utilized different DNAm arrays; we examined approximately 30 times more CpG sites and haplogroup H or J carriers. However, the DNAm profiles in peripheral blood do not necessarily reflect similar methylation changes in other tissues (30).

The meta-analysis showed no evidence of prediabetes-specific heterogeneity. Further studies using more homogeneous cohorts or larger sample sizes should be conducted to gain more insight into the interplay between mtDNA variation and DNAm in the setting of prediabetes.

The maternal inheritance of mtDNA could create male–female asymmetry in the consequences of mtDNA mutations since mtSNPs that only affect males will not be subject to natural selection (31). This hypothesis has been tested in *Drosophila melanogaster* in which a strong effect of mtSNPs on gene expression was observed only in males, while the mitochondrial effect in females was negligible (32). In humans, there is no evidence of sex-specific mitochondrial genetic control of peripheral blood gene expression (5,6). Even though sex-specific DNAm patterns have been demonstrated in peripheral blood (33,34), our results imply that, similarly to gene expression, mtDNA variation has the same genetic effect on peripheral blood DNAm in both sexes.

Strengths and limitations

The present study has strengths and limitations that warrant consideration. The variants in the YFS were obtained through next-generation sequencing, which allowed us to study a broad range of mtDNA variants. Genotyping a part of the LURIC participants with two different microarrays increased the quality of haplogroup assignment. We were also able to verify the self-reported smoking status with the cotinine measurements in the LURIC study. The discovery analyses were adjusted for bias and inflation using a state-of-the-art method that was specifically developed for epigenome-wide association studies, which maximizes power while properly controlling the false-positive rate (35). Still, it is important to note that, as with any (epi)genome-wide association study, it is possible that some of the identified associations represent false positives.

The main weakness was the lack of a comparable mtDNA data set for replication, as many of the sequenced mtSNPs in the YFS were not genotyped in the LURIC. For some mtSNPs, replication was sought by using a tagged mtSNP, which could have resulted in false-positive or false-negative replications. The smoking status in the YFS was only self-reported and was not verified by cotinine measurements. Finally, it should be highlighted that the YFS is a population-based study, whereas the LURIC participant pool mainly consists of older patients referred to coronary angiography. As DNAm variation has been associated with age numerous times, the difference in the ages between the participants of the two cohorts may have affected the results. Also, other confounding factors owing to contrasting participant characteristics may have yielded an effect on the results since socioeconomic status (36) and lipid composition (37), for instance, have independent effects on DNAm.

Conclusion

This study provides evidence of a mitochondrial genetic control of autosomal DNAm, with little evidence found for sex- and prediabetes-specific effects. The functional relevance of the identified associations remains unclear. Further replication studies, preferably using sequencing data and more homogeneous study groups, should be conducted to thoroughly establish the mitochondrial genetic determinants of DNAm.

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

Materials and Methods

Study populations

The YFS (http://youngfinnsstudy.utu.fi) is a Finnish longitudinal population study on the evolution of cardiovascular risk factors from childhood to adulthood (38). We utilized data from two follow-ups conducted in 2007 and 2011, including 2204 and 2060 participants, respectively. Phenotypic information and DNAm data were collected in 2011, and mtDNA data were obtained from the 2007 follow-up samples. The study was approved by the ethics committee of the Hospital District of Southwest Finland, and the study protocol of each study phase corresponded with the proposal of the World Health Organization.

The LURIC study consists of 3316 patients of German ancestry who underwent coronary angiography between 1997 and 2000 at a tertiary care center in Southwestern Germany (39). The clinical indications for angiography were chest pain or non-invasive tests that were consistent with myocardial ischemia. Patients with any acute illness other than acute coronary syndrome, any predominant non-cardiac disease and/or a history of malignancy within the past 5 years were excluded from the study. The study plan was approved by the ethics committee of the State Chamber of Physicians of Rhineland-Palatinate.

All participants in both cohorts gave their written informed consent, and the studies were conducted in accordance with the Declaration of Helsinki.

DNAm assessment and quality control

In both cohorts, genomic DNA was extracted from peripheral whole blood samples by standardized methods. DNAm levels were quantified using the Illumina Infinium MethylationEPIC BeadChip according to the manufacturer's protocols. The array covers over 850 000 methylation sites across the nuclear DNA.

In the YFS, DNAm data were processed using the minfi Bioconductor package in R (40). All analyzed samples had a sum of detection *P*-values across all probes of <0.05. The log2 -median of methylated and unmethylated intensities among the analyzed samples clustered within the default threshold (10.5) of the getQC function in minfi. Samples for which the self-reported sex did not match with the predicted sex obtained with the getSex function in minfi were excluded. Background subtraction and dye-bias normalization were performed via the noob method (41), and stratified quantile normalization was performed using the preprocessQuantile function, both implemented in minfi. Probes with a detection *P*-value of >0.01 in 99% of the samples and cross-reactive probes (42,43) were excluded from the analysis. Probes with SNPs were removed using the dropLociwithSnps function in minfi. After quality control, the total number of autosomal CpGs was 769 683 in 1529 samples. In addition, the sex-specific analyses included 17 334 X-chromosomal CpGs.

In the LURIC study, quality control was implemented using the CPACOR pipeline (44), excluding samples with a call rate of ≤ 0.95 and those that showed sex discordance. CpGs located in close proximity (1–2 bp) to a genetic polymorphism in the European population with a frequency of >0.01% as well as cross-reactive probes and probes with a detection *P*-value of >0.05 in at least 1% of the samples were removed using the rmSNPandCH function in the DMRcate package (45), followed by quantile normalization. A total of 795 619 autosomal and 18 138 X-chromosomal CpGs from 2423 samples were included in further analyses.

Beta values [ranging between 0 (no methylation) and 1 (full methylation)] were calculated according to the equation $b = M/(M + U + 100)$, where M and U denote the methylated and the unmethylated signals, respectively.

mtDNA sequencing and data processing in the YFS

In the YFS, mtSNPs were determined by next-generation sequencing. The pipeline has been described in detail earlier (46). In brief, mtDNA was amplified from genomic DNA samples ($n = 1807$) and was sequenced with the Illumina HiSeq system. Reads from all samples that achieved any mean bait coverage ($n = 1658$) were aligned with the revised Cambridge Reference Sequence (1) and were analyzed using Mutserve version 1.2.1, a stand-alone version of the web tool mtDNA-Server (47), with the default parameters. Variants overlapping with any mtDNA-like sequence in the nucleus (NUMTs) were excluded. The list of NUMTs insertions was based on the work by Dayama *et al.* (48). The minimum heteroplasmy level was set at 0.05—we defined sites with a heteroplasmy level below this threshold as homoplasmic wild-type alleles and those with a heteroplasmy level >0.95 as homoplasmic variants. Mutserve identified variants in 1365 different nucleotide positions from 1657 samples. We required each sequenced sample to have an overall mean coverage of ≥30 and 1531 samples survived this threshold. The average coverage across all samples was 536.

Samples without complete phenotype, DNAm and mtDNA data were excluded after which 926 samples (525 women and 401 men) remained for further analysis, with 241 mtSNPs having an allele frequency of ≥0.01. Heteroplasmic variants were excluded owing to their low number. To reduce the computational effort, we selected a set of 37 tagging mtSNPs (Supplementary Material, Table S2) that captured 126 other mtSNPs with a linkage disequilibrium of $r^2 \geq 0.8$ by using Tagger (49) and HaploView (50). Seventy-eight mtSNPs were not tagged by any another variant, which resulted in 115 mtSNPs to be included in the association analyses.

Haplogroups were determined by using HaploGrep version 2.2.0 (51) (Phylotree build 17) (52). For association testing, the haplogroups were assigned to major haplogroups. Haplogroups with a frequency of <0.01 and samples whose haplogroup quality score was <0.90 were excluded, leaving 863 samples with nine haplogroups for the haplogroup–CpG analysis.

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

mtDNA genotyping and data processing in the LURIC study

Genomic DNA was extracted from peripheral blood, and the mtSNPs were genotyped using the Illumina HumanExome-12 version 1.2 BeadChip ($n = 1981$) and the Illumina 200 k MetaboChip ($n = 3150$) microarrays. Samples with a call rate of <0.95, sex mismatch and cryptic relatedness (pi-hat > 0.2) were removed using PLINK version 1.90b6.21. Variants with an allele frequency of <0.005 and a call rate of <0.95 were also excluded. Heterozygous genotypes possibly owing to mitochondrial heteroplasmy were coded as missing.

After quality control and the exclusion of samples with missing phenotype, DNAm or mtDNA data, 1456 and 2290 samples from the HumanExome-12 and MetaboChip arrays, respectively, were available for further analyses. Of the variants genotyped with these arrays, 53 HumanExome-12 mtSNPs and 42 MetaboChip mtSNPs had an allele frequency of ≥0.005. Most of the genotyped individuals ($n = 1429$) were present in both arrays, and the total number of individuals was 2317 (718 women and 1599 men).

Haplogroups were assigned by applying HaploGrep separately to the two genotyping batches by applying the '–chip' parameter. We included haplogroups based on two criteria: (1) a quality score ≥0.90 in at least one genotyping batch, or (2) a quality score of ≥0.80 and the same major haplogroup assigned in both arrays. This resulted in 998 samples to be included in the haplogroup–CpG analysis.

Definition of clinical variables

Height and weight were measured, and body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared. Sex was self-reported. In the YFS, the smoking history of the participants was self-reported and was classified into six categories based on smoking frequency (active smoker or at least once a day, once a week or more often but not daily, less often than once a week, attempts to quit, has quit and has never smoked). In the LURIC study, smoking status was also self-reported but was additionally verified by the measurement of serum cotinine concentration. A commonly used cut-off to define active smoking is 15 μg/l (53), and we used this value to reclassify self-reported non- or ex-smokers as active smokers. Participants were categorized into five groups: heavy smokers (defined as smoking ≥20 cigarettes per day), light smokers, former smokers who quit smoking <10 years ago, former smokers who quit smoking >10 years ago and never-smokers.

The classification of prediabetes was based on the criteria of the American Diabetes Association (54). Venous blood samples were drawn after an overnight fast for the determination of serum glucose and glycated hemoglobin A_{1c} (HbA1c). Individuals with prediabetes were defined as having a fasting plasma glucose (FPG) level of 5.6–6.9 mmol/l, a 2-h plasma glucose level of 7.8–11.0 mmol/l during a 75-g oral glucose tolerance test (OGTT), or an HbA1c level of 39–47 mmol/mol without a diagnosis of T2D. The diagnosis of T2D included individuals with an FPG level of ≥7.0 mmol/l, a 2-h glucose level of ≥11.1 mmol/l during an OGTT or an HbA1c level of ≥48 mmol/mol, or those who reported using oral glucose-lowering medication or insulin (but had not reported having type 1 diabetes) or who reported having been diagnosed with T2D by a physician. Those diagnosed with type 1 diabetes were also ruled out. OGTTs were performed only for the LURIC participants.

Discovery analysis in the YFS

Differentially methylated CpG loci for mtSNPs were identified using a linear regression model where the methylation beta values were modeled as a linear function of the presence (coded as 1) or absence (coded as 0) of the variant allele using the lm function in R. The model involved adjustment for age, sex, BMI, smoking status, white blood cell type proportions, methylation batch effects and principal components (PCs) derived from the mtDNA data. The fraction of white blood cells (CD8T, CD4T, NK cells, B cells, monocytes and granulocytes) was estimated through the reference-based Houseman method (55) using the estimateCellCounts function in the minfi package (40). Methylation batch affects were addressed by including the first five PCs of array control probes in the regression models. PC analysis was performed on all mtDNA genotypes that passed quality control using the logisticPCA package (56). The use of mitochondrial PCs as covariates has been demonstrated to be a robust method to adjust for population stratification in genetic association studies. In addition, the use of mitochondrial PCs effectively removes false-positive associations but does not cause a loss of power in detecting true associations (57,58). All CpG–mtSNP analyses were adjusted for the first six mitochondrial PCs. CpG loci were considered differentially methylated if they reached a Bonferroni-corrected *P*-value of 7.8×10^{-10} ($9 \times 10^{-8}/115$) based on the number of independent tests in a whole blood EPIC array (59) and the number of mtSNPs.

Differential methylation between haplogroups

We applied a similar linear model to flag CpG sites for those showing differential methylation between the nine haplogroups. We selected the most common haplogroup H to be the reference to which other haplogroups were compared. Mitochondrial PCs were excluded from the covariates since the haplogroups are strongly correlated with the mitochondrial PCs. Significance was defined as $P < 1.0 \times 10^{-8}$ ($9 \times 10^{-8}/9$).

Sex- and prediabetes-specific analyses

The sex-specific effects of mtSNPs on methylation beta values were tested by applying the same linear model as described above to males and females separately. Differences in effect sizes were compared by applying a fixed-effect inverse variance-weighted meta-analysis model to

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

each CpG–mtSNP pair by pooling the effect estimates and standard errors from males and females in Genome-Wide Association Meta-Analysis (GWAMA) software version 2.1 (60). Heterogeneity was examined by calculating the sex-heterogeneity *P*-value (61). A significant *P*-value suggests that there is a significant difference in effect sizes between the sexes. A minimum variant allele count of 10 in both sexes was required, which resulted in 63 mtSNPs to be included and a significance threshold of 1.4×10^{-9} ($9 \times 10^{-8}/63$).

The effect of prediabetes on the association between mtSNPs and DNAm was studied similarly by applying the linear model separately to individuals with prediabetes and normoglycemic controls and by pooling the results in GWAMA. The number of mtSNPs was 47, and significance was defined as $P < 1.9 \times 10^{-9}$ ($9 \times 10^{-8}/47$).

Control for bias and inflation

We corrected the effect estimates, their standard errors and the corresponding *P*-values for bias and inflation using the R package bacon (35), and all reported results are bacon-corrected. We used the inflation function in the same package to compute the inflation factor λ for each association analysis from all CpG–mtSNP/haplogroup pairs that were analyzed. The regime of minimal inflation is $\lambda < 1.14$ (35).

Replication in the LURIC study

We sought replication in the LURIC study by applying the same linear regression models as in the discovery phase. We included variants with an allele frequency of >0.005, or a minimum variant allele count of five in the sex- and prediabetes-specific analyses. If a tagging mtSNP from the discovery sample was not genotyped in the replication sample, an mtSNP for replication was searched from the tagged mtSNPs. If several tagged mtSNPs were genotyped, linear regression was performed on all tagged variants and the sentinel mtSNP with the smallest association *P*-value was used.

Associations were considered fully replicated if the replication *P*-value from the linear regression model fell below a Bonferroni-corrected *P*-value of $0.05/n$, with n being the number of significant associations in the discovery study covered in the replication sample. For nominal replication, the *P*-value threshold was set at 0.05. We also required consistent effect directions across both cohorts and in males/females and individuals with/without prediabetes. The two Illumina microarrays were analyzed separately, including the 12 overlapping mtSNPs present in both arrays, thus providing the opportunity of validation in the case of significant results. Associations with $P < 0.05$ in one genotyping batch and with $P > 0.05$ in another batch were not regarded as replications.

We benchmarked the observed replication rates for general mtSNP and haplogroup analyses by calculating the expected degree of replication. First, we used a false-discovery rate inverse quantile transformation to correct the effect sizes for the winner's curse (62) and also took into account the lower number of mtSNPs available in the replication cohort. Second, we calculated the expected number of associations meeting the Bonferroni-corrected replication threshold by using the method described in Okbay *et al*. (63)

Finally, we performed a fixed-effect inverse variance-weighted meta-analysis of the replicated associations by combining the effect estimates and standard errors from the discovery and replication cohorts with the GWAMA software. An association was considered to be significant if the meta-analysis *P*-value fell below the significance threshold used in the corresponding discovery analysis. The inverse variance-based method compensates for the varying number of samples in the cohorts by allowing larger studies to have more weight in the analysis (64).

Expression quantitative trait methylation analysis

To gain insight into whether our association data were connected to biological processes, we examined the associations between peripheral blood genome-wide transcriptomics and the differentially methylated CpG sites identified in the replication phase. Gene expression and DNAm data were available for 1364 YFS participants. The expression data were analyzed using the Illumina HumanHT-12 v4 Expression BeadChip. The procedures have been described previously (6).

CpGs were regressed against cell count proportions and the first 30 PCs of the array control probes. Similarly, the 19 637 transcription probes were regressed against the first 20 PCs derived from the expression data. For each CpG site, expression probes within a 2 Mb window (± 1 Mb) were included. Linear regression was applied between the residuals from the CpG regression (explanatory variable) and the expression probe residuals (dependent variable). The model was additionally adjusted for age, sex and BMI. The *P*-value for statistical significance was defined as 0.05 divided by the number of combinations between CpGs and genes.

Supplementary Material

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

Data availability

The data sets generated and analyzed during the current study comprise health-related participant data, and their use is therefore restricted under the regulations concerning professional secrecy (Act on the Openness of Government Activities, 612/1999) and sensitive personal data (Personal Data Act, 523/1999, implementing the EU data protection directive 95/46/EC). Owing to these legal constraints, the individual-level data cannot be stored in public repositories or otherwise made publicly available but are, however, available from the authors upon a reasonable request.

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

Funding

This work was supported by personal grants awarded to J.L. by the Juhani Aho Foundation, Laboratoriolääketieteen edistämissäätiö, the Aarne Koskelo Foundation and the Finnish Medical Foundation. The YFS has been financially supported by the Academy of Finland (grants 322098 to T.L.; 330809 and 338395 to E.R.; 286284 and 134309 [Eye]; 126925, 121584, 124282 and 129378 [Salve]; 117787 [Gendi] and 41071 [Skidi]); the Kela Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility Area of Kuopio, Tampere and Turku University Hospitals (grant X51001); the Juho Vainio Foundation; the Paavo Nurmi Foundation; the Finnish Foundation for Cardiovascular Research; the Finnish Cultural Foundation; the Sigrid Juselius Foundation; the Tampere Tuberculosis Foundation; the Emil Aaltonen Foundation; the Yrjö Jahnsson Foundation; the Signe and Ane Gyllenberg Foundation; the Diabetes Research Foundation of the Finnish Diabetes Association; the EU Horizon 2020 programme (grant 848146 for To_Aition; grant 755320 for TAXINOMISIS); the European Research Council (grant 742927 for MULTIEPIGEN project); the Tampere University Hospital Supporting Foundation and the Finnish Society of Clinical Chemistry. The genome-wide methylation analyses in the LURIC cohort were financially supported by the German Federal Ministry for Education and Research (BMBF; grant agreement numbers 01EA1808A and 01EA1411A) within the framework of the German Competence Cluster for Nutrition and Cardiovascular Health (nutriCARD) (Halle-Jena-Leipzig).

References

1. 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. *Nat. Genet.*, **23**, 147.
2. Chinnery, P.F. and Hudson, G. (2013) Mitochondrial genetics. *Br. Med. Bull.*, **106**, 135–159.
3. Matilainen, O., Quirós, P.M. and Auwerx, J. (2017) Mitochondria and epigenetics–crosstalk in homeostasis and stress. *Trends Cell Biol.*, **27**, 453–463.
4. Lopes, A.F.C. (2020) Mitochondrial metabolism and DNA methylation: a review of the interaction between two genomes. *Clin. Epigenetics*, **12**, 182.
5. Kassam, I., Qi, T., Lloyd-Jones, L., Holloway, A., Bonder, M.J., Henders, A.K., Martin, N.G., Powell, J.E., Franke, L., Montgomery, G.W. *et al.* (2016) Evidence for mitochondrial genetic control of autosomal gene expression. *Hum. Mol. Genet.*, **25**, 5332–5338.
6. Laaksonen, J., Seppälä, I., Raitoharju, E., Mononen, N., Lyytikäinen, L.-P., Waldenberger, M., Illig, T., Lepistö, M., Almusa, H., Ellonen, P. *et al.* (2019) Discovery of mitochondrial DNA variants associated with genome-wide blood cell gene expression: a population-based mtDNA sequencing study. *Hum. Mol. Genet.*, **28**, 1381–1391.
7. Atilano, S.R., Malik, D., Chwa, M., Cáceres-del-Carpio, J., Nesburn, A.B., Boyer, D.S., Kuppermann, B.D., Jazwinski, S.M., Miceli, M.V., Wallace, D.C. *et al.* (2015) Mitochondrial DNA variants can mediate methylation status of inflammation, angiogenesis and signaling genes. *Hum. Mol. Genet.*, **24**, 4491–4503.
8. Bellizzi, D., D'Aquila, P., Giordano, M., Montesanto, A. and Passarino, G. (2012) Global DNA methylation levels are modulated by mitochondrial DNA variants. *Epigenomics*, **4**, 17–27.
9. Min, J.L., Hemani, G., Hannon, E., Dekkers, K.F., Castillo-Fernandez, J., Luijk, R., Carnero-Montoro, E., Lawson, D.J., Burrows, K., Suderman, M. *et al.* (2021) Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat. Genet.*, **53**, 1311–1321.
10. Huan, T., Joehanes, R., Song, C., Peng, F., Guo, Y., Mendelson, M., Yao, C., Liu, C., Ma, J., Richard, M. *et al.* (2019) Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nat. Commun.*, **10**, 4267.
11. Lemire, M., Zaidi, S.H.E., Ban, M., Ge, B., Aïssi, D., Germain, M., Kassam, I., Wang, M., Zanke, B.W., Gagnon, F. *et al.* (2015) Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. *Nat. Commun.*, **6**, 6326.
12. Walaszczyk, E., Luijten, M., Spijkerman, A.M.W., Bonder, M.J., Lutgers, H.L., Snieder, H., Wolffenbuttel, B.H.R. and van Vliet-Ostaptchouk, J.V. (2018) DNA methylation markers associated with type 2 diabetes, fasting glucose and HbA 1c levels: a systematic review and replication in a case-control sample of the Lifelines study. *Diabetologia*, **61**, 354–368.
13. Davegårdh, C., García-Calzón, S., Bacos, K. and Ling, C. (2018) DNA methylation in the pathogenesis of type 2 diabetes in humans. *Mol. Metab.*, **14**, 12–25.
14. Matsha, T.E., Pheiffer, C., Humphries, S.E., Gamieldien, J., Erasmus, R.T. and Kengne, A.P. (2016) Genome-wide DNA methylation in mixed ancestry individuals with diabetes and prediabetes from South Africa. *Int. J. Endocrinol.*, **2016**, 1–11.
15. Chinnery, P.F., Elliott, H.R., Patel, S., Lambert, C., Keers, S.M., Durham, S.E., McCarthy, M.I., Hitman, G.A., Hattersley, A.T. and Walker, M. (2005) Role of the mitochondrial DNA 16184-16193 poly-C tract in type 2 diabetes. *Lancet*, **366**, 1650–1651.
16. Saxena, R., De Bakker, P.I.W., Singer, K., Mootha, V., Burtt, N., Hirschhorn, J.N., Gaudet, D., Isomaa, B., Daly, M.J., Groop, L. *et al.* (2006) Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am. J. Hum. Genet.*, **79**, 54–61.
17. Chinnery, P.F., Mowbray, C., Patel, S.K., Elson, J.L., Sampson, M., Hitman, G.A., McCarthy, M.I., Hattersley, A.T. and Walker, M. (2007) Mitochondrial DNA haplogroups and type 2 diabetes: a study of 897 cases and 1010 controls. *J. Med. Genet.*, **44**, e80–e80.
18. Achilli, A., Olivieri, A., Pala, M., Kashani, B.H., Carossa, V., Perego, U.A., Gandini, F., Santoro, A., Battaglia, V., Grugni, V. *et al.* (2011) Mitochondrial DNA backgrounds might modulate diabetes complications rather than T2DM as a whole. *PLoS One*, **6**, e21029.
19. West, A.P. and Shadel, G.S. (2017) Mitochondrial DNA in innate immune responses and inflammatory pathology. *Nat. Rev. Immunol.*, **17**, 363–375.
20. Oshiumi, H., Matsumoto, M., Hatakeyama, S. and Seya, T. (2009) Riplet/RNF135, a RING finger protein, ubiquitinates RIG-I to promote interferon-β induction during the early phase of viral infection∗. *J. Biol. Chem.*, **284**, 807–817.
21. Dor, Y. and Cedar, H. (2018) Principles of DNA methylation and their implications for biology and medicine. *Lancet*, **392**, 777–786.
22. Martínez-Reyes, I. and Chandel, N.S. (2020) Mitochondrial TCA cycle metabolites control physiology and disease. *Nat. Commun.*, **11**, 102.

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

23. Wiese, M. and Bannister, A.J. (2020) Two genomes, one cell: mitochondrial-nuclear coordination via epigenetic pathways. *Mol. Metab.*, **38**, 100942.
24. Kopinski, P.K., Janssen, K.A., Schaefer, P.M., Trefely, S., Perry, C.E., Potluri, P., Tintos-Hernandez, J.A., Singh, L.N., Karch, K.R., Campbell, S.L. *et al.* (2019) Regulation of nuclear epigenome by mitochondrial DNA heteroplasmy. *PNAS*, **116**, 16028–16035.
25. Lee, W.T., Sun, X., Tsai, T.-S., Johnson, J.L., Gould, J.A., Garama, D.J., Gough, D.J., McKenzie, M., Trounce, I.A. and John, J.C.S. (2017) Mitochondrial DNA haplotypes induce differential patterns of DNA methylation that result in differential chromosomal gene expression patterns. *Cell Death Dis.*, **3**, 1–11.
26. Marbaix, A.Y., Tyteca, D., Niehaus, T.D., Hanson, A.D., Linster, C.L. and Van Schaftingen, E. (2014) Occurrence and subcellular distribution of the NAD(P)HX repair system in mammals. *Biochem. J.*, **460**, 49–58.
27. Rydström, J. (2006) Mitochondrial NADPH, transhydrogenase and disease. *Biochim. Biophys. Acta Bioenerg.*, **1757**, 721–726.
28. Tretter, L. and Adam-Vizi, V. (2005) Alpha-ketoglutarate dehydrogenase: a target and generator of oxidative stress. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **360**, 2335–2345.
29. Cortés-Pereira, E., Fernández-Tajes, J., Fernández-Moreno, M., Vázquez-Mosquera, M.E., Relaño, S., Ramos-Louro, P., Durán-Sotuela, A., Dalmao-Fernández, A., Oreiro, N., Blanco, F.J. *et al.* (2019) Differential association of mitochondrial DNA haplogroups J and H with the methylation status of articular cartilage: potential role in apoptosis and metabolic and developmental processes. *Arthritis Rheum.*, **71**, 1191–1200.
30. Langevin, S.M. and Kelsey, K.T. (2013) The fate is not always written in the genes: Epigenomics in epidemiologic studies. *Environ. Mol. Mutagen.*, **54**, 533–541.
31. 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 Ecol. Evol.*, **19**, 238–244.
32. Innocenti, P., Morrow, E.H. and Dowling, D.K. (2011) Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. *Science*, **332**, 845–848.
33. Singmann, P., Shem-Tov, D., Wahl, S., Grallert, H., Fiorito, G., Shin, S.Y., Schramm, K., Wolf, P., Kunze, S., Baran, Y. *et al.* (2015) Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenetics Chromatin*, **8**, 43–43.
34. McCarthy, N.S., Melton, P.E., Cadby, G., Yazar, S., Franchina, M., Moses, E.K., Mackey, D.A. and Hewitt, A.W. (2014) Meta-analysis of human methylation data for evidence of sex-specific autosomal patterns. *BMC Genom.*, **15**, 981–981.
35. van Iterson, M., van Zwet, E.W., Heijmans, B.T. and the BIOS Consortium (2017) Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.*, **18**, 19.
36. Fiorito, G., McCrory, C., Robinson, O., Carmeli, C., Rosales, C.O., Zhang, Y., Colicino, E., Dugué, P.-A., Artaud, F., McKay, G.J. *et al.* (2019) Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: a multi-cohort analysis. *Aging*, **11**, 2045–2070.
37. Gomez-Alonso, M.D.C., Kretschmer, A., Wilson, R., Pfeiffer, L., Karhunen, V., Seppälä, I., Zhang, W., Mittelstraß, K., Wahl, S., Matias-Garcia, P.R. *et al.* (2021) DNA methylation and lipid metabolism: an EWAS of 226 metabolic measures. *Clin. Epigenetics*, **13**, 7.
38. Juonala, M., Viikari, J.S. and Raitakari, O.T. (2013) Main findings from the prospective cardiovascular risk in young Finns study. *Curr. Opin. Lipidol.*, **24**, 57–64.
39. Winkelmann, B.R., März, W., Boehm, B.O., Zotz, R., Hager, J., Hellstern, P. and Senges, J. (2001) Rationale and design of the LURIC study—a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics*, **2**, S1–S73.
40. Fortin, J.-P., Triche, T.J., Jr. and Hansen, K.D. (2017) Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics*, **33**, 558–560.
41. Triche, T.J., Jr., Weisenberger, D.J., Van Den Berg, D., Laird, P.W. and Siegmund, K.D. (2013) Low-level processing of Illumina Infinium DNA methylation bead arrays. *Nucleic Acids Res.*, **41**, e90–e90.
42. Pidsley, R., Zotenko, E., Peters, T.J., Lawrence, M.G., Risbridger, G.P., Molloy, P., Van Djik, S., Muhlhausler, B., Stirzaker, C. and Clark, S.J. (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.*, **17**, 208.
43. McCartney, D.L., Walker, R.M., Morris, S.W., McIntosh, A.M., Porteous, D.J. and Evans, K.L. (2016) Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data*, **9**, 22–24.
44. Lehne, B., Drong, A.W., Loh, M., Zhang, W., Scott, W.R., Tan, S.-T., Afzal, U., Scott, J., Jarvelin, M.-R., Elliott, P. *et al.* (2015) A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.*, **16**, 37.
45. Peters, T.J., Buckley, M.J., Statham, A.L., Pidsley, R., Samaras, K., V Lord, R., Clark, S.J. and Molloy, P.L. (2015) De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin*, **8**, 6.
46. Laaksonen, J., Mishra, P.P., Seppälä, I., Lyytikäinen, L.-P., Raitoharju, E., Mononen, N., Lepistö, M., Almusa, H., Ellonen, P., Hutri-Kähönen, N. *et al.* (2021) Examining the effect of mitochondrial DNA variants on blood pressure in two Finnish cohorts. *Sci. Rep.*, **11**, 611.
47. Weissensteiner, H., Forer, L., Fuchsberger, C., Schöpf, B., Kloss-Brandstätter, A., Specht, G., Kronenberg, F. and Schönherr, S. (2016) mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Research*, **44**, W64–W69.
48. 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**, 12640–12649.
49. de Bakker, P.I.W., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. and Altshuler, D. (2005) Efficiency and power in genetic association studies. *Nat. Genet.*, **37**, 1217–1223.
50. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
51. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A. and Schönherr, S. (2016) Haplo Grep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.*, **44**, W58–W63.
52. van Oven, M. (2015) PhyloTree Build 17: growing the human mitochondrial DNA tree. *Forensic Sci. Int. Genet. Suppl. Ser.*, **5**, e392–e394.
53. Benowitz, N.L., Bernert, J.T., Foulds, J., Hecht, S.S., Jacob, P., III, Jarvis, M.J., Joseph, A., Oncken, C. and Piper, M.E. (2020) Biochemical verification of tobacco use and abstinence: 2019 update. *Nicotine Tob. Res.*, **22**, 1086–1097.

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

54. American Diabetes Association (2014) Diagnosis and classification of diabetes mellitus. *Diabetes Care*, **37**, S81–S90.
55. 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 Bioinform.*, **13**, 86.
56. Landgraf, A.J. and Lee, Y. (2020) Dimensionality reduction for binary data through the projection of natural parameters. *J. Multivar. Anal.*, **180**, 104668.
57. Biffi, A., Anderson, C.D., Nalls, M.A., Rahman, R., Sonni, A., Cortellini, L., Rost, N.S., Matarin, M., Hernandez, D.G., Plourde, A. *et al.* (2010) Principal-component analysis for assessment of population stratification in mitochondrial medical genetics. *Am. J. Hum. Genet.*, **86**, 904–917.
58. Miller, B., Arpawong, T.E., Jiao, H., Kim, S.-J., Yen, K., Mehta, H.H., Wan, J., Carpten, J.C. and Cohen, P. (2019) Comparing the utility of mitochondrial and nuclear DNA to adjust for genetic ancestry in association studies. *Cell*, **8**, 306.
59. Mansell, G., Gorrie-Stone, T.J., Bao, Y., Kumari, M., Schalkwyk, L.S., Mill, J. and Hannon, E. (2019) Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genom.*, **20**, 366.
60. Mägi, R. and Morris, A.P. (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinform.*, **11**, 288–288.
61. Mägi, R., Lindgren, C.M. and Morris, A.P. (2010) Meta-analysis of sex-specific genome-wide association studies. *Genet. Epidemiol.*, **34**, 846–853.
62. Bigdeli, T.B., Lee, D., Webb, B.T., Riley, B.P., Vladimirov, V.I., Fanous, A.H., Kendler, K.S. and Bacanu, S.-A. (2016) A simple yet accurate correction for winner's curse can predict signals discovered in much larger genome scans. *Bioinformatics*, **32**, 2598–2603.
63. Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S.F.W. *et al.* (2016) Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, **533**, 539–542.
64. Haidich, A.B. (2010) Meta-analysis in medical research. *Hippokratia*, **14**, 29–37.

Downloaded from https://academic.oup.com/hmg/article/31/10/1720/6431953 by guest on 21 June 2022

PUBLICATION

III

Examining the effect of mitochondrial DNA variants on blood pressure in two Finnish cohorts

Laaksonen, J., Mishra, P. P., Seppälä, I., Lyytikäinen, L.-P., Raitoharju, E., Mononen, N., Lepistö, M., Almusa, H., Ellonen, P., Hutri-Kähönen, N., Juonala, M., Raitakari, O., Kähönen, M., Salonen, J. T. and Lehtimäki, T.

Scientific Reports 11 (2021), 611

Publication reprinted with the permission of the copyright holders

scientific reports

OPEN

Examining the effect of mitochondrial DNA variants on blood pressure in two Finnish cohorts

Jaakko Laaksonen[1]✉, Pashupati P. Mishra[1], Ilkka Seppälä[1], Leo-Pekka Lyytikäinen[1], Emma Raitoharju[1], Nina Mononen[1], Maija Lepistö[2], Henrikki Almusa[2], Pekka Ellonen[2], Nina Hutri-Kähönen[3], Markus Juonala[4,5,6], Olli Raitakari[7,8,9], Mika Kähönen[10], Jukka T. Salonen[11,12] & Terho Lehtimäki[1]

High blood pressure (BP) is a major risk factor for many noncommunicable diseases. The effect of mitochondrial DNA single-nucleotide polymorphisms (mtSNPs) on BP is less known than that of nuclear SNPs. We investigated the mitochondrial genetic determinants of systolic, diastolic, and mean arterial BP. MtSNPs were determined from peripheral blood by sequencing or with genome-wide association study SNP arrays in two independent Finnish cohorts, the Young Finns Study and the Finnish Cardiovascular Study, respectively. In total, over 4200 individuals were included. The effects of individual common mtSNPs, with an additional focus on sex-specificity, and aggregates of rare mtSNPs grouped by mitochondrial genes were evaluated by meta-analysis of linear regression and a sequence kernel association test, respectively. We accounted for the predicted pathogenicity of the rare variants within protein-encoding and the tRNA regions. In the meta-analysis of 87 common mtSNPs, we did not observe significant associations with any of the BP traits. Sex-specific and rare-variant analyses did not pinpoint any significant associations either. Our results are in agreement with several previous studies suggesting that mtDNA variation does not have a significant role in the regulation of BP. Future studies might need to reconsider the mechanisms thought to link mtDNA with hypertension.

High blood pressure (BP) is a major public health problem. It is an established risk factor for many noncommunicable diseases, such as cardiovascular disease, renal dysfunction, and dementia, and causes over 9 million premature deaths globally per year[1]. It is generally accepted that the regulation of BP is a multi-factorial trait, involving lifestyle, environmental, and genetic factors. To date, the majority of the genetic variants have been identified in studies of the nuclear genome[2–4], and a limited number of studies have explicitly investigated the associations with variation in the mitochondrial genome. This gap in genetic knowledge is of a particular interest because, in addition to cellular energy production, mitochondria modulate, for example, the intracellular dynamics of nitric oxide, reactive oxygen species, and Ca^{2+}, which, in turn, control endothelial function in blood vessels[5].

[1]Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Arvo Ylpön katu 34, PO Box 100, 33014 Tampere, Finland. [2]Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland. [3]Department of Paediatrics, Tampere University Hospital and Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [4]Department of Medicine, University of Turku, Turku, Finland. [5]Division of Medicine, Turku University Hospital, Turku, Finland. [6]Murdoch Children's Research Institute, Parkville, VIC, Australia. [7]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland. [8]Research Centre for Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [9]Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland. [10]Department of Clinical Physiology, Tampere University Hospital and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [11]Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland. [12]MAS-Metabolic Analytical Services Oy, Helsinki, Finland. ✉email: jaakko.h.laaksonen@tuni.fi

Mitochondrial DNA (mtDNA) is a maternally inherited, double-stranded circular molecule containing 16,569 nucleotides that encode 22 transfer RNAs (tRNAs), two ribosomal RNAs (rRNAs), and 13 protein subunits of the four oxidative phosphorylation (OXPHOS) complexes[6]. The majority of the proteins functioning in the mitochondria are, however, encoded in nuclear DNA. There are multiple copies of mtDNA within each cell, and the mtDNA mutation rate is significantly higher than that of nuclear DNA[7]. Given the high mutation rate and the fact that both common and rare variants may influence the disease phenotypes[8], the most appropriate method to study mtDNA is through sequencing. One possible mechanism through which mutations in the mtDNA contribute to BP variation is the oxidative stress due to an increased production of reactive oxygen species, which in turn causes cardiovascular and renal damage[9–11]. In addition to mutations in the mtDNA, the mitochondrial dysfunction may also be the consequence of mutations in the nuclear-encoded mitochondrial genes, or result from environmental and lifestyle factors.

Several studies with varying ethnic groups have demonstrated that a mutational hot spot for mitochondrial single-nucleotide polymorphisms (mtSNPs) associated with hypertension is in the tRNA-encoding genes[10,12,13]. However, many of these studies have been conducted on related individuals, and it is likely that most of these hypertension-associated variants are inherited and rare on a population level. While the role of naturally occurring mtDNA variation is still incompletely understood[14], some evidence for the association of mtSNPs with BP exists from well-established cohorts. In the Framingham Heart Study, one mtSNP in OXPHOS complex IV correlated with systolic blood pressure (SBP)[15]. In a recent two-cohort sequencing study of older North American adults, both common and rare mtSNPs in the tRNA region were associated with variation in SBP in white participants[16]. Negative results have also been reported: 64 tagging mtSNPs that efficiently capture common mtDNA variation in the European population were not associated with BP in a study consisting of over 2000 individuals[17], and a sequencing study conducted in South African population which utilized the MutPred pathogenicity prediction scores did not find a significant role for mtDNA variation in association with blood pressure levels[18]. A lexical tree analysis with phylogenetically related mtDNA variants in European population identified significant relationships with some common diseases, e. g. multiple sclerosis, but not with hypertension[19].

The goal of the present study was to investigate the effect of mtDNA variants on SBP, diastolic blood pressure (DBP), and mean arterial pressure (MAP) among participants of two independent Finnish cohorts. We sought to both discover new and replicate the previously reported mitochondrial genetic determinants of BP. We studied both the effects of common single mtSNPs and the pooled effects of rare variants across seven mtDNA regions.

Materials and methods

Study participants. The Young Finns Study (YFS, http://youngfinnsstudy.utu.fi) is a Finnish longitudinal population study on the evolution of cardiovascular risk factors from childhood to adulthood[20]. We used the phenotype data from the follow-up in 2011 which included 2060 participants in total. The blood samples for mtDNA sequencing were obtained during the 2007 follow-up, when 2204 participants were examined. The study was approved by the ethical committee of the Hospital District of Southwest Finland (ETMK:68/1801/2017), and the study protocol of each study phase corresponded with the proposal by the World Health Organization.

The Finnish Cardiovascular Study (FINCAVAS) participant pool consists of patients recruited during 2001–2007 who underwent exercise stress tests at Tampere University Hospital[21]. A total of 4,068 participants completed a technically successful exercise test. The main indications for the exercise test were a suspicion of coronary heart disease (frequency 46%), the evaluation of work capacity (26%), testing for vulnerability to arrhythmia during exercise (25%), and assessing the adequacy of coronary heart disease treatment (13%); some patients had more than one indication. The study protocol was approved by the Ethical Committee of the Pirkanmaa Hospital District, Finland (R00153).

All participants in both study cohorts gave their written informed consent, and the studies were conducted in accordance with the Declaration of Helsinki.

Blood pressure and other clinical measurements. In the YFS, BP was determined as the average of three measurements taken at two-minute intervals in a sitting position from the right arm brachial artery with a random zero sphygmomanometer. Korotkoff's first phase was used as the sign of SBP and the fifth phase as the sign of DBP. In the FINCAVAS, the patients lay in the supine position for 10 min, after which BP was measured once by an experienced nurse using a brachial cuff according to the Korotkoff's method.

In both cohorts, the observed BPs were adjusted for antihypertensive medication usage. The medications were self-reported by the study participants, the duration of treatment was not known, and adherence was not assessed. Adjusted SBP was calculated by increasing the recorded measure by 8, 14, and 20 mmHg for 1, 2, and ≥ 3 medication classes taken, respectively. DBP measurements were adjusted similarly by increasing the recorded measure by 4, 10, and 16 mmHg for 1, 2, and ≥ 3 medication classes taken, respectively[22]. This adjustment method has proven to work with both simulated and real-life data[23], and it maximizes the genetic and shared environmental variance components while minimizing individual-specific components[22]. FINCAVAS participants displayed clearly more antihypertensive medication usage (31%, 24%, and 16% on 1, 2, and ≥ 3 medication classes taken, respectively) than YFS participants (7%, 3%, and 0.3%, respectively). The mean adjustment was 10/6 and 13/9 mmHg for treated YFS and FINCAVAS participants, respectively. Adjusted MAP was calculated from the adjusted SBP and DBP values as $MAP = DBP + 0.333 \times (SBP—DBP)$.

Height and weight were measured, and body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared.

MtDNA sequencing and quality control in the YFS. Genomic DNA sample (n = 1807) concentrations were measured from whole-blood samples with the Qubit BR dsDNA kit (Life Technologies). MtDNA

was amplified from the genomic DNA using the REPLI-g mtDNA kit (Qiagen) in a 50 µl reaction volume. The primer sites in the REPLI-g kit have been previously validated[24]. After the amplification mtDNA samples were processed into Illumina sequencing-compatible libraries with the Nextera DNA sample preparation kit (Illumina). The mtDNA concentrations were measured with Qubit dsDNA for Nextera tagmentation reaction. The reaction volume in the Nextera tagmentation and amplification steps was 20 µl, and after both steps the libraries were purified with an EdgeBio Performa V3 96-Well Short Plate (Edge BioSystems). After the amplification, the libraries were first incubated with 4 µl of EdgeBio SOPE resin and then purified with EdgeBio Performa plates. After purification, 48 samples with different index tags were pooled together (2 µl each) in each pool and concentrated with DNA Clean & Concentrator-5 (Zymo Research). The final volume of the concentrated pool was 15 µl. The sequencing-ready libraries were quantitated with an Agilent 2100 Bioanalyzer High Sensitivity kit (Agilent). The libraries were sequenced with the Illumina HiSeq system. All samples (n = 1658) that achieved any mean bait coverage were included in the further processing steps.

Paired-end FASTQ files were aligned with the revised Cambridge Reference Sequence (rCRS)[6] using the sequence aligner BWA-MEM v. 0.7.17[25]. Reads mapped to the rCRS were sorted according to the start position and written into a BAM file using SAMtools v. 1.8[26]. Homo- and heteroplasmic variants were then detected with Mutserve v. 1.2.1, a stand-alone version of the web tool mtDNA-server[27]. We applied the default thresholds for mapping, base, and alignment quality scores in the Phred scale: 20, 20, and 30, respectively. Mutations in the mtDNA can affect either all or a varying proportion of the mtDNA molecules, and the terms used for these phenomena are homoplasmy and heteroplasmy, respectively[7]. The minimum heteroplasmy level was set to 0.05—we defined sites with a heteroplasmy level below this threshold as homoplasmic wild-type alleles and sites with a heteroplasmy level above 0.95 as homoplasmic variants. Mutserve applies a Bayesian model to the detection of homoplasmic variants, and in order to call a wild-type allele, we required a minimum sequencing coverage of five on both the forward and reverse strands. Mutserve identified mtSNPs in 1,365 different nucleotide positions from 1657 samples. Mean coverages per sample and per mtSNP were 497 and 525, respectively. With this coverage, the 0.05 heteroplasmy detection level is quite reliable[28]. Genotypes for heteroplasmic variants overlapping with any mtDNA-like sequence in the nucleus (NUMTs) were set to missing. The list of NUMTs insertions was based on the work by Dayama et al.[29] We required each sequenced sample to have an overall mean coverage of ≥ 50, and 1,434 samples had a mean coverage above this threshold. Then, samples that did not have both phenotype and mtDNA data available were excluded, after which 1,150 samples remained for further analysis. After this, mean coverages per sample and per mtSNP were 563 and 568, respectively. Of the 1,365 identified common and rare mtSNPs, 249 had an allele frequency of ≥ 0.01. Heteroplasmy level between 0.05 and 0.95 in at least 1% of the samples was observed only for one mtSNP, m.16192C>T in the hypervariable region. This mtSNP was not genotyped in the FINCAVAS arrays and hence not included in the meta-analysis. In general, heteroplasmy rate was low in most sites, for 99% of the sites the number of heteroplasmic samples was three or less. This finding is in line with a previous study which found the incidence of heteroplasmy to be higher in tissues with high metabolic activity[30].

MtDNA genotyping and quality control in the FINCAVAS. Genomic DNA was extracted from peripheral blood leukocytes by using the QIAamp DNA Blood Minikit and automated biorobot M48 extraction (Qiagen). We applied the Illumina Cardio-MetaboChip and HumanCoreExome-12 v1.1 SNP arrays for genotyping mtSNPs. Genotyping was completed for 2,824 and 1,032 samples using the Cardio-MetaboChip and HumanCoreExome arrays, respectively. Genotypes were called using Illumina's GenomeStudio GenCall algorithm. Samples with call rate of < 0.95, excess heterozygosity, cryptic relatedness (pi-hat > 0.2) and sex mismatch, as well as genetic outliers based on multi-dimensional scaling (MDS) plots, were removed. mtSNPs with a call rate of < 0.95 and a Hardy–Weinberg equilibrium p-value of $\leq 10^{-6}$ were also removed. Heterozygous genotypes possibly due to mitochondrial heteroplasmy or technical error were coded as missing. Homozygosity at a genotyped mtSNP indicates genotype frequency close to zero or one. After quality control, 53 mtSNPs from 2273 samples genotyped with the Cardio-MetaboChip array and 146 mtSNPs from 926 samples genotyped with the HumanCoreExome array were available. For four individuals that were genotyped with both arrays, the genotypes for the 34 overlapping mtSNPs were set as missing in the Cardio-MetaboChip array. After including only samples with both phenotype and mtDNA data available, 2193 and 923 samples from the Cardio-MetaboChip and HumanCoreExome arrays, respectively, were available for association analyses. The total number of individuals was 3112 since four samples were genotyped with both arrays. Of the variants genotyped with these arrays, 36 Cardio-MetaboChip mtSNPs and 67 HumanCoreExome mtSNPs had an allele frequency of ≥ 0.01.

Statistical analyses. *Association analyses of common variants.* In order to investigate the associations of SBP, DBP, and MAP with mtDNA variants, BP levels were modeled as a linear function of the presence (coded as 1) or absence (coded as 0) of the variant allele using the lm function in R. Heteroplasmic genotypes were set to missing. In order to achieve normality of the BP distributions, we applied the rescaled inverse normal transformation, i.e. multiplied the inverse normal transformed BP values by the standard deviation of the original trait values in each cohort. This strategy makes the distributions normal and controls the type I error, and restores the original scale of measurement and thus enhances the power of meta-analysis[31]. Inverse normal transformation also effectively deals with phenotypic outliers[32]. Age, sex, and BMI were added as covariates to the linear regression models and the p-values were calculated using a standard F test with one degree of freedom. Association analyses were performed separately for the three data sets, and a random-effect meta-analysis was then performed using GWAMA software[33]. Variants with an allele frequency of ≥ 0.01 were included. The total number of mtSNPs included in the meta-analysis was 87, of which 22 were present in YFS and in both FINCAVAS data sets. The number of mtSNPs present in YFS and only in the FINCAVAS participants genotyped with

the HumanCoreExome array was 52, and 13 mtSNPs were present in YFS and only in participants genotyped with the Cardio-MetaboChip array. Using Matrix Spectral Decomposition (matSpDlite)[34] with the method of Li and Ji[35], we determined that 45 of the 87 mtSNPs represented an estimate of the number of independent genetic effects for mtDNA. This resulted in a Bonferroni-corrected significance level of 0.0011 (i.e. 0.05/45). In the correction for multiple testing, we did not account for the testing of three BP traits, since they were correlated (Pearson's correlation coefficients between adjusted SBP, DBP, and MAP values 0.78–0.96 in the YFS and 0.67–0.92 in FINCAVAS).

Study power. With a sample size of ~ 2000, meaning that an mtSNP with an allele frequency of ≥ 0.01 was present in the YFS and the smaller FINCAVAS data set, our single-variant analysis had ~ 80% power to detect an mtSNP that explains 2.5% of the variance in BP. When an mtSNP with an allele frequency of ≥ 0.01 was present in the YFS and both FINCAVAS data sets, our analysis had ~ 95% and ~ 80% power to detect mtSNPs explaining 2.5% and 1% of the variance, respectively[36].

Sex-specific analyses. The power of sex-combined analysis (i.e. males and females analyzed together) is reduced when heterogeneity is present in the allelic effects between the sexes. We examined the possible heterogeneity by applying the same linear model as described above to males and females separately, and by applying sex-differentiated meta-analysis in GWAMA[33,37]. In total, 66 mtSNPs with an allele frequency of ≥ 0.01 in both sexes were included in the meta-analysis. A sex-differentiated *p*-value below the individual *p*-values for males and females is indicative of a significant association with both sexes. We also tested for sex-specific heterogeneity, which is equivalent to a formal test of interaction with sex. A significant heterogeneity *p*-value would suggest that there is a difference in the effect sizes between the sexes. By using the same Matrix Spectral Decompositon as above, we now determined that 33 of the 66 mtSNPs represented an estimate of the number of independent genetic effects for mtDNA. To account for the two sexes tested, the significance level was now defined as $0.05/33/2 = 7.6 \times 10^{-4}$.

Rare-variant analyses. Standard methods used to test for common variant associations are underpowered for detecting associations with rare variants[38]. The power of haplogroup-based analysis may also be insufficient if potential causative variants at a same mtDNA site are scattered across divergent haplogroups. The most common approaches for the analysis of rare-variant associations are the burden test and the sequence kernel association test (SKAT)[38,39]. The former collapses the rare variants within a specified region, assuming that all variants are either deleterious or protective. The latter aggregates and tests the collective effects of rare variants within a region without assuming similar directionality or effect size for each variant. Therefore, SKAT is superior to burden test for analyzing regions where both risk and protective variants as well as noncausal variants may be present[39,40] and it has been successfully applied for both sequenced and genotyped mtDNA data[16,41,42].

We used Mitomap[43] to cluster the variants into seven regions, including each of the four OXPHOS complexes (I, III, IV, and V), all rRNAs combined, all tRNAs combined, and control region and non-coding regions combined. Homoplasmic alleles were coded as 0 or 2, corresponding to the reference and variant allele, respectively. Heteroplasmic genotypes were introduced in a dosage matrix, similarly to imputed genotypes. The dosage was calculated as twice the heteroplasmy rate. We employed a SKAT meta-analysis with minor allele frequencies (MAFs) of ≤ 0.01 (T1) and ≤ 0.05 (T5) for SBP, DBP, and MAP, using the seqMeta package in R, with the default beta weights, and with age, sex, and BMI as covariates. Bonferroni-corrected statistical significance was defined as 0.0071 (i.e. 0.05/7).

In order to assess the functional relevance of non-synonymous variants, we predicted the pathogenicity of the identified variants with MutPred[44,45] and MitoTIP[46] pathogenicity scores. The MutPred algorithm assigns each variant in the protein-encoding mtDNA regions a pathogenicity score between 0 and 1. Variants with a MutPred score > 0.5 are potentially "harmful", and variants with a score > 0.75 should be considered a high confidence "harmful". MitoTIP predicts the pathogenicity of the variants in the tRNA regions, and the prediction scores have been interpreted within quartiles as "likely benign", "possibly benign", "possibly pathogenic" or "likely pathogenic". We leveraged SKAT meta-analysis for the BP levels, similarly as described above, but including only variants with a MutPred score > 0.5 or a MitoTIP classification "possibly" or "likely pathogenic".

Control for population stratification. The use of mtDNA principal components (PCs) as covariates has been demonstrated to be a robust method to adjust for population stratification in genetic association studies. In addition, the use of mitochondrial PCs effectively removes false-positive associations but does not cause a loss of power in detecting true associations[47,48]. Logistic PC analysis was performed on all homoplasmic genotypes passing quality control and with a MAF of ≥ 0.01 using the logisticPCA package[49]. For each data set, we selected the number of mitochondrial PCs for the single variant analyses so that the median chi-squared-based genomic inflation factor (λ_{GC}) was as close to one as possible. For SKAT, we selected the same number of PCs as were used in the sex-combined common variant analyses. Values of $\lambda_{GC} < 1.05$ are generally considered benign[50]. For any result data sets with $\lambda_{GC} > 1.05$, genomic control was applied by multiplying the standard errors of regression coefficients by the square root of the inflation factor of the respective study.

Results

On average, participants in the FINCAVAS were older and had higher BMI and BP levels than participants in the YFS (Table 1). We evaluated the associations of 87 common mtDNA variants with SBP, DBP, and MAP using random-effect meta-analysis in these two Finnish cohorts, with sample sizes of up to 4,262. When both sexes were analyzed together, we did not observe any statistically significant associations. The top three associations

	YFS n = 1150	FINCAVAS n = 3112
Age, years	42.2 (5.0)	56.3 (13.0)
Women, n (%)	661 (57.5)	1,214 (39.0)
Body mass index, kg/m^2	26.5 (5.00)	27.5 (4.5)
Unadjusted SBP, mmHg	118.7 (13.9)	136.7 (18.9)
Unadjusted DBP, mmHg	74.8 (10.4)	80.0 (9.6)
Unadjusted MAP, mmHg	89.5 (10.8)	98.9 (11.3)
Medication-adjusted SBP, mmHg	119.7 (14.7)	145.6 (21.1)
Medication-adjusted DBP, mmHg	75.4 (10.8)	86.2 (11.2)
Medication-adjusted MAP, mmHg	90.2 (11.4)	106.0 (13.3)

Table 1. Baseline characteristics of the YFS and FINCAVAS cohorts. Values are means (SD). *YFS* Young Finns Study, *FINCAVAS* Finnish Cardiovascular Study, *SBP* systolic blood pressure, *DBP* diastolic blood pressure, *MAP* mean arterial pressure.

mtSNP	Trait	Beta (SE)	*p*	MAF	n
m.1243T>C	SBP	− 4.22 (1.79)	0.019	0.024	4,219
m.15257G>A	MAP	4.91 (2.20)	0.025	0.013	2,071
m.11674C>T	SBP	− 4.26 (1.91)	0.026	0.029	3,342

Table 2. Three most significant associations in the sex-combined meta-analysis. *mtSNP* mitochondrial single-nucleotide polymorphism, *Trait* the trait used for a specific test, *Beta (SE)* beta coefficient and the corresponding standard error, *p* unadjusted *p*-value from meta-analysis, *MAF* minor allele frequency, *n* sample size contributing in a particular mtSNP meta-analysis.

are shown in Table 2. All meta-analysis results and the quantile–quantile plots from the cohort-level analyses are provided in the Supplemental material, in Table S1 and Fig. S1.

There was no evidence for differences between the sexes in the mitochondrial genetic control of BP levels. All results from the sex-differentiated and heterogeneity meta-analysis as well as the quantile–quantile plots from the cohort-level analyses are provided in the Supplemental material, in Table S2 and Figs. S2–S3.

Finally, we conducted SKAT meta-analyses on all rare (T1 test) and low-frequency (T5 test) variants. We also employed SKAT taking into account the predicted pathogenicity of the variants. None of the analyses yielded significant associations with the BP traits over of the tested mtDNA regions. All results from SKAT meta-analysis are provided in the Supplemental material, Tables S3 and S4.

Discussion

We conducted one the largest studies to date investigating the possible mitochondrial genetic determinants of BP. We did not find any associations that survived after correction for multiple testing.

We could not replicate the results where the common mtSNPs m.3197T>C and m.15924A>G were associated with higher SBP and MAP, respectively, in white North American individuals[16]. In the current study, the former mtSNP was sequenced or genotyped in over 2000 and the latter in over 4200 individuals, which casts doubt over the role of these two mtSNPs in the variation of BP. Possible confounding effects due to different genetic ancestry should not have had an effect on the associations, since both the previous and the current study included mitochondrial PCs as regression covariates[47,48]. However, it should be noted that, since the North American individuals were significantly older than the Finns in our cohorts, it is possible that the regulative role of these two mtSNPs is activated only in later life.

The aforementioned North American study[16] identified significant pooled effects on SBP across variants in the tRNA regions in white participants. Our analysis of rare and low-frequency variants did not yield any significant associations, not even when we accounted for the predicted pathogenicity of the variants within protein-encoding and the tRNA regions. We did not account for the functional relevance of the variants in the rRNA regions because to the best of our knowledge, no tools exist for annotating pathogenicity for variants in those regions.

Another mtSNP previously found to be associated with SBP is m.5913G>A in the Framingham Heart Study[15]. In the current study, this mtSNP was only sequenced in the YFS population with a MAF of < 0.01, while it was not genotyped in FINCAVAS, which left us unable to study the effect of this mtSNP in single-variant analysis. While the sample size of the current study is, to the authors' knowledge, one of largest used in investigating mtDNA associations with BP, our inability to replicate earlier findings from well-characterized cohorts underlines the need for mtDNA association studies with significantly larger sample sizes from multiple cohorts. For example, a recent study including ~ 170,000 individuals from 45 cohorts reported associations with seven metabolic outcomes, but these did not include BP[42]. These large consortium studies can however be extremely costly, and another basis for future research could be achieved by introducing more homogeneous study groups with less confounding effects. We should also be ready to accept to the null hypothesis that mtDNA variants do not

contribute to BP variation on a significant level as it is implied by several previous studies[17–19]. Another hypothesis to be tested is that instead of being causal of BP variation, the mtDNA variants would impact the hypertension complications or alter the course of the disease[51].

The strength of the current study was that the mtSNPs in the YFS were obtained through sequencing, which allowed us to study the effects of both common and rare variants. Another strength was the large range of blood pressure variation in our two independent cohorts, and the method we used to adjust for antihypertensive treatment effects has proven to work across a wire variety of clinical scenarios[22,23]. Some limitations should also be acknowledged—the mtSNPs in FINCAVAS were genotyped, resulting in a relatively small number of mtSNPs included in the meta-analyses. The majority of the rare mtSNPs were identified only in the YFS population which decreased the power of the SKAT meta-analysis. We also applied a stringent threshold to define homoplasmic variants and wild-type alleles, but relaxing the threshold to e. g. 10% and 90% would have increased the number of homoplasmic alleles only marginally. Another limitation is that we only investigated substitutions, whereas mtDNA deletions and insertions were not identified. In addition, BP measurements were performed only once for the FINCAVAS participants. Ambulatory BP monitoring could greatly increase the robustness of the future studies.

In conclusion, we found no support for the hypothesis that common or rare mtSNPs play a significant role in the regulation of BP. While studies with larger sample sizes might show different results, we should also be open to the idea that the non-significant outcomes reported by the current and previous studies are in fact correct, and that future studies concerning this topic need to reconsider the mechanisms thought to link mtDNA with hypertension.

Data availability

The data sets generated and analyzed during the current study comprise health-related participant data, and their use is therefore restricted under the regulations concerning professional secrecy (Act on the Openness of Government Activities, 612/1999) and sensitive personal data (Personal Data Act, 523/1999, implementing the EU data protection directive 95/46/EC). Due to these legal constraints, the Ethics Committee of the Hospital District of Southwest Finland has stated in 2016 that individual-level data cannot be stored in public repositories or otherwise made publicly available. Data are, however, available from the authors upon a reasonable request.

Received: 11 October 2020; Accepted: 10 December 2020
Published online: 12 January 2021

References

1. Lim, S. S. *et al.* A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2224–2260 (2012).
2. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* **41**, 666–676 (2009).
3. Ehret, G. B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
4. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat. Genet.* **48**, 1162–1170 (2016).
5. Davidson, S. M. & Duchen, M. R. Endothelial mitochondria: Contributing to vascular function and disease. *Circ. Res.* **100**, 1128–1141 (2007).
6. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
7. Chinnery, P. F. & Hudson, G. Mitochondrial genetics. *Br. Med. Bull.* **106**, 135–159 (2013).
8. Schork, N. J., Murray, S. S., Frazer, K. A. & Topol, E. J. Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* **19**, 212–219 (2009).
9. Dikalov, S. I. & Ungvari, Z. Role of mitochondrial oxidative stress in hypertension. *Am. J. Physiol. Heart Circ.* **305**, 1417–1427 (2013).
10. Ding, Y., Xia, B., Yu, J., Leng, J. & Huang, J. Mitochondrial DNA mutations and essential hypertension (review). *Int. J. Mol. Med.* **32**, 768–774 (2013).
11. Lahera, V., De Las Heras, N., López-Farré, A., Manucha, W. & Ferder, L. Role of mitochondrial dysfunction in hypertension and obesity. *Curr. Hypertens. Rep.* **19**, 11 (2017).
12. Wilson, F. H. *et al.* A cluster of metabolic defects caused by mutation in a mitochondrial DNA. *Science* **306**, 1190–1194 (2004).
13. Wang, S. *et al.* Maternally inherited essential hypertension is associated with the novel 4263A>G mutation in the mitochondrial tRNA Ile gene in a Large Han Chinese Family. *Circ. Res.* **108**, 862–870 (2011).
14. Bray, A. W. & Ballinger, S. W. Mitochondrial DNA mutations and cardiovascular disease. *Curr. Opin. Cardiol.* **32**, 267–274 (2017).
15. Liu, C. *et al.* Association of genetic variation in the mitochondrial genome with blood pressure and metabolic traits. *Hypertension* **60**, 949–956 (2012).
16. Buford, T. W. *et al.* Mitochondrial DNA sequence variants associated with blood pressure among 2 cohorts of older adults. *J. Am. Heart Assoc.* **7**, e010009. https://doi.org/10.1161/JAHA.118.010009 (2018).
17. Saxena, R. *et al.* Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am. J. Hum. Genet.* **79**, 54–61 (2006).
18. Venter, M., Malan, L., van Dyk, E., Elson, J. L. & van der Westhuizen, F. H. Using MutPred derived mtDNA load scores to evaluate mtDNA variation in hypertension and diabetes in a two-population cohort: The SABPA study. *J. Genet. Genomics* **44**, 139–149 (2017).
19. Hudson, G., Gomez-Duran, A., Wilson, I. J. & Chinnery, P. F. Recent mitochondrial DNA mutations increase the risk of developing common late-onset human diseases. *PLOS Genet.* **10**, e1004369. https://doi.org/10.1371/journal.pgen.1004369 (2014).
20. Juonala, M., Viikari, J. S. & Raitakari, O. T. Main findings from the prospective cardiovascular risk in Young Finns Study. *Curr. Opin. Lipidol.* **24**, 57–64 (2013).
21. Nieminen, T. *et al.* The Finnish Cardiovascular Study (FINCAVAS): Characterising patients with high risk of cardiovascular morbidity and mortality. *BMC Cardiovasc. Disord.* **6**, 9. https://doi.org/10.1186/1471-2261-6-9 (2006).
22. Cui, J. S., Hopper, J. L. & Harrap, S. B. Antihypertensive treatments obscure familial contributions to blood pressure variation. *Hypertension* **41**, 207–210 (2003).

23. Tobin, M. D., Sheehan, N. A., Scurrah, K. J. & Burton, P. R. Adjusting for treatment effects in studies of quantitative traits: Antihypertensive therapy and systolic blood pressure. *Stat. Med.* **24**, 2911–2935 (2005).
24. Laaksonen, J. *et al.* Discovery of mitochondrial DNA variants associated with genome-wide blood cell gene expression: A population-based mtDNA sequencing study. *Hum. Mol. Genet.* **28**, 1381–1391 (2019).
25. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
26. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
27. Weissensteiner, H. *et al.* mtDNA-Server: Next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Res.* **44**, W64–W69 (2016).
28. Weissensteiner, H. *et al.* Haplocheck: Phylogeny-based contamination detection in mitochondrial and whole-genome sequencing studies. *biorxiv* https://doi.org/10.1101/2020.05.06.080952v2(2020) (2020).
29. Dayama, G., Emery, S. B., Kidd, J. M. & Mills, R. E. The genomic landscape of polymorphic human nuclear mitochondrial insertions. *Nucleic Acids Res.* **42**, 12640–12649 (2014).
30. Naue, J. *et al.* Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. *Mitochondrion* **20**, 82–94 (2015).
31. Tang, Z. Z. & Lin, D. Y. Meta-analysis for discovering rare-variant associations: Statistical methods and software programs. *Am. J. Hum. Genet.* **97**, 35–53 (2015).
32. Auer, P. L., Reiner, A. P. & Leal, S. M. The effect of phenotypic outliers and non-normality on rare-variant association testing. *Eur. J. Hum. Genet.* **24**, 1188–1194 (2016).
33. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinform.* **11**, 288. https://doi.org/10.1186/1471-2105-11-288 (2010).
34. Nyholt, D. R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74**, 765–769 (2004).
35. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221–227 (2005).
36. McRae, A. F., Byrne, E. M., Zhao, Z., Montgomery, G. W. & Visscher, P. M. Power and SNP tagging in whole mitochondrial genome association studies. *Genome Res.* **18**, 911–917 (2008).
37. Mägi, R., Lindgren, C. M. & Morris, A. P. Meta-analysis of sex-specific genome-wide association studies. *Genet. Epidemiol.* **34**, 846–853 (2010).
38. Li, B. & Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* **83**, 311–321 (2008).
39. Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence Kernel association test. *Am. J. Hum. Genet.* **89**, 82–93 (2011).
40. Basu, S. & Pan, W. Comparison of statistical tests for disease association with rare variants. *Genet. Epidemiol.* **35**, 606–619 (2011).
41. Giorgi, E. E. *et al.* No association between the mitochondrial genome and prostate cancer risk: The multiethnic cohort. *Cancer Epidemiol. Biomarkers Prev.* **25**, 1001–1003 (2016).
42. Kraja, A. T. *et al.* Associations of mitochondrial and nuclear mitochondrial variants and genes with seven metabolic traits. *Am. J. Hum. Genet.* **104**, 112–138 (2019).
43. Lott, M. T. *et al.* mtDNA variation and analysis using mitomap and mitomaster. *Curr. Protoc. Bioinform.* **44**, 23–26 (2013).
44. Mort, M. *et al.* In silico functional profiling of human disease-associated and polymorphic amino acid substitutions. *Hum. Mutat.* **31**, 335–346 (2010).
45. Pereira, L., Soares, P., Radivojac, P., Li, B. & Samuels, D. C. Comparing phylogeny and the predicted pathogenicity of protein variations reveals equal purifying selection across the global human mtDNA diversity. *Am. J. Hum. Genet.* **88**, 433–439 (2011).
46. Sonney, S. *et al.* Predicting the pathogenicity of novel variants in mitochondrial tRNA with MitoTIP. *PLOS Comput. Biol.* **13**, e1005867. https://doi.org/10.1371/journal.pcbi.1005867 (2017).
47. Biffi, A. *et al.* Principal-component analysis for assessment of population stratification in mitochondrial medical genetics. *Am. J. Hum. Genet.* **86**, 904–917 (2010).
48. Miller, B. *et al.* Comparing the utility of mitochondrial and nuclear DNA to adjust for genetic ancestry in association studies. *Cells* **8**, 306. https://doi.org/10.3390/cells8040306 (2019).
49. Landgraf, A. J. & Lee, Y. Dimensionality reduction for binary data through the projection of natural parameters. *J. Multivar. Anal.* **180**, 104668. https://doi.org/10.1016/j.jmva.2020.104668 (2020).
50. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459–463 (2010).
51. Achilli, A. *et al.* Mitochondrial DNA backgrounds might modulate diabetes complications rather than T2DM as a whole. *PLoS ONE* **6**, e21029. https://doi.org/10.1371/journal.pone.0021029 (2011).

Acknowledgements

This work was supported by personal Grants awarded to JL by the Juhani Aho Foundation, Laboratoriolääketieteen edistämissäätiö, the Aarne Koskelo Foundation, and the Finnish Medical Foundation. The Young Finns Study (YFS) has been financially supported by the Academy of Finland: Grants 322098, 286284, 134309 [Eye], 126925, 121584, 124282, 129378 [Salve], 117787 [Gendi], and 41071 [Skidi]; the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (Grant X51001); the Juho Vainio Foundation; the Paavo Nurmi Foundation; the Finnish Foundation for Cardiovascular Research; the Finnish Cultural Foundation; the Sigrid Juselius Foundation; the Tampere Tuberculosis Foundation; the Emil Aaltonen Foundation; the Yrjö Jahnsson Foundation; the Signe and Ane Gyllenberg Foundation; the Diabetes Research Foundation of Finnish Diabetes Association; the EU Horizon 2020 program (Grant 755320 for TAXINOMISIS; Grant 848146 for To_Aition); the European Research Council (Grant 742927 for MULTIEPIGEN project); the Tampere University Hospital Supporting Foundation; and the Finnish Society of Clinical Chemistry. The Finnish Cardiovascular Study (FINCAVAS) has been financially supported by the Competitive Research Funding of Tampere University Hospital (Grant 9M048 and 9N035), the Finnish Cultural Foundation, the Finnish Foundation for Cardiovascular Research, the Emil Aaltonen Foundation, the Tampere Tuberculosis Foundation, the EU Horizon 2020 program (Grant 755320 for TAXINOMISIS; Grant 848146 for To_Aition), and the Academy of Finland Grant 322098. In addition, we thank Jari Kaikkonen and Janne Asikainen for their technical assistance.

Author contributions

J.L. designed the study, conducted statistical analyses and wrote the main manuscript text. P.M., I.S. and L.P.L. contributed to the statistical analyses. E.R., N.M., M.L., H.A. and P.E. contributed to data collection. N.H.K., M.J.

and M.K. participated in cohort collection. O.R. handled funding and participated in cohort collection. J.S. contributed to data collection. T.L. supervised the study, handled funding, participated in cohort collection, in addition to reviewing and editing the manuscript. All authors have contributed to and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-79931-6.

Correspondence and requests for materials should be addressed to J.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021