# IDENTIFYING CIRCULATING MEDIATORS OF CEREBROVASCULAR DISEASE

# IDENTIFYING CIRCULATING MEDIATORS OF CEREBROVASCULAR DISEASE

By MICHAEL R. CHONG, B.ASc (Hons.), MSc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2021) Hamilton, Ontario, Canada.

(Biochemistry)

TITLE: Identifying circulating mediators of cerebrovascular disease

AUTHOR: Michael R. Chong BASc, MSc (McMaster University)

SUPERVISOR: Dr. Guillaume Paré, MD, MSc

NUMBER OF PAGES: xx, 256

## LAY ABSTRACT

Current stroke medications work by targeting circulating molecules. Our aim was to discover new drug candidates by combining genetic and circulating biomarker data using a technique called “Mendelian Randomization”. In Study 1, we screened 653 circulating proteins and found evidence supporting causal roles for two novel candidates, SCARA5 and TNFSF12. Prior experimental studies suggest an important role for mitochondria in stroke recovery. Accordingly, in Study 2, we characterized the genetic basis of an emerging biomarker, mitochondrial DNA copy number (mtDNA-CN). Analyses of 395,781 participants revealed 71 associated genetic regions, representing a 40% increase in our knowledge. In Study 3, we measured mtDNA-CN in 3,498 acute patients and observed that lower levels predicted elevated risk of worse post-stroke functional outcomes. Furthermore, Mendelian Randomization analysis suggested a likely causal relationship. Overall, this work uncovered several novel therapeutic leads for preventing stroke onset and progression that warrant further investigation to verify therapeutic utility.

**ABSTRACT**

Many current drugs for stroke act by targeting circulating molecules, yet these have not been exhaustively evaluated for therapeutic potential. A central challenge is that while many molecules correlate with stroke risk, only a subset cause stroke. To disentangle causality from association, a statistical genetics framework called "Mendelian Randomization" can be used by integrating genetic, biomarker, and phenotypic information. In Study 1, we screened 653 circulating proteins using this technique and found evidence supporting causal roles for seven proteins, two of which (SCARA5 and TNFSF12) were not previously implicated in stroke pathogenesis. We also characterized potential side-effects of targeting these molecules for stroke prevention and did not identify any adverse effects for SCARA5. The remaining two studies focused on investigating the role of an emerging marker of mitochondrial activity, leukocyte mitochondrial DNA copy number (mtDNA-CN). Mitochondria have long been known to play a protective role in stroke recovery; however, a mitochondrial basis for stroke protection has not been extensively studied in humans. In Study 2, we first sought to better understand the genetic basis of mtDNA-CN in a series of genetic association studies involving 395,781 UK residents. We identified 71 loci which represents a 40% increase in our knowledge. In Study 3, epidemiological analyses of 3,498 acute stroke demonstrated that low mtDNA-CN was associated with higher risk of subsequent mortality and worse functional outcome 1-month after stroke. Furthermore, Mendelian Randomization analyses corroborated a causative relationship for the first time, implying that interventions that increase mtDNA-CN levels in stroke patients may represent a novel strategy for mitigating post-stroke complications. Ultimately, this work uncovered several novel therapeutic leads for preventing stroke onset and ameliorating its progression. Future investigations are necessary to better understand the underlying biological mechanisms connecting these molecules to stroke and to further interrogate their validity as potential drug targets.

**ACKNOWLEDGEMENTS**

First, I would like to extend my gratitude to my family. To my partner, Adrienne: thank you for your endless support. While research can be a solitary endeavour, we went through the lows and highs together – from frustrating nights of code debugging to the elation of manuscript acceptance. To my parents: thank you for being constant beacons of wisdom. Your food care packages allowed me to focus on my work, and home-made butter tarts distracted and "buttered up" my committee members. To my grandparents: thank you for instilling in your grandsons the primacy of education, which inspired us to become a teacher, veterinarian, and researcher. To my brothers: thank you for keeping me grounded and for reading my papers (or at least pretend to).

Second, I would like to thank my committee members, Drs. Andrew McArthur and Ashkan Shoamanesh. Andrew: thank you for asking the tough questions, giving me opportunities to lecture, and being a role model bioinformatician; your leadership in the genomic surveillance of COVID-19 is truly inspiring. Ashkan: thank you for being my friendly work neighbour, answering my naïve questions about stroke, and showing me that old guys can still hoop.

Finally, I would like to thank the Genetic and Molecular Epidemiology Laboratory (GMEL) who have been like a second family to me. To all of my GMEL colleagues: your friendship, guidance, and the moments we shared together defined my graduate student experience. A special thanks to the best supervisor and mentor anyone could ask for, Dr. Guillaume Paré. You showed me by example what it takes to be an exceptional researcher – dedication, perseverance, and hard work. You gave me the confidence and encouragement to pursue my own ideas, which showed me the value of having the freedom to explore one's own curiosities. Above all else, you have taught me countless life lessons beyond research that I am forever grateful for. Merci beaucoup!

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF ABBREVIATIONS

ABC – age, biomarker, and clinical history score
ABO – Histo-blood group ABO system transferase
APOE – apolipoprotein E
ATP – adenosine triphosphate
AutoMitoC – automatic mitochondrial DNA copy number estimation pipeline
B2M – beta 2 microglobulin
CATIS – China Antihypertensive Trial in Acute Ischemic Stroke
CCL2 – chemokine ligand 2
CD40 – cluster of differentiation 40
CD163 – cluster of differentiation 163
CHI3L1 – chitinase-3-like protein 1
CI – confidence interval
COMPASS – Cardiovascular Outcomes for People using Anticoagulation Strategies
CoVasc-ICH – Colchicine for the prevention of vascular events after an acute IntraCerebral Hemorrhage
COX1 – D Cyclooxygenase
CRP – C-reactive protein
CST3 – cystatin C
CT – computed tomography
DEPICT – Data-driven Expression-Prioritized Integration for Complex Traits
DGUOK – deoxyguanosine kinase
DNA – deoxyribonucleic acid
dNTP – deoxynucleoside triphosphate
ENCODE – Encyclopedia of DNA elements
eQTL – expression quantitative trait loci
ExWAS – exome-wide association study
F11 – coagulation factor XI
FDR – false discovery rate
Fn14 – fibroblast growth factor inducible-14
GDF15 – growth differentiation factor 15
GISCOME – Genetics of Ischemic Stroke funCtional Outcome study
GMEL – Genetic and Molecular Epidemiological Laboratory
GTPBP3 – mitochondrial GTP binding protein 3
GWAS – genome-wide association study
HR – Hazard Ratio
IL1 – Interleukin 1
IL6 – Interleukin 6
IMPROVE – multicenter, longitudinal carotid intima-media thickness and IMT progression as predictors of vascular events in a high-risk European population
INTERSTROKE – Importance of Conventional and Emerging Risk Factors of Stroke in Different Regions and Ethnic Groups of the World
ISGC – International Stroke Genetics Consortium

KB – kilobases
KORA F4 – cooperative health research in the region of Augsburg F4
L2R – $\log_2$ relative ratio
LD – linkage disequilibrium
LDLR – low-density lipoprotein receptor
LONP1 – mitochondrial lon peptidase 1
LPA or LP(a) – lipoprotein(a)
LRPPRC – leucine rich pentatricopeptide repeat containing
MGME1 – mitochondrial genome maintenance exonuclease 1
MIEF1 – mitochondrial elongation factor 1
MMP9 – matrix metalloproteinase 9
MMP12 – matrix metalloproteinase 12
MR – Mendelian randomization
MRI – magnetic resonance imaging
MR-PRESSO – Mendelian Randomization Pleiotropy RESidual Sum and Outlier
mRS – modified Rankin Scale
MRPS35 – mitochondrial ribosomal protein S35
MT – mitochondrial
mtDNA – mitochondrial DNA
mtDNA-CN – mtDNA copy number
NAD – nicotinamide dinucleotide
NAMPT – nicotinamide phosphribosyltransferase
Nfl – Neurofilament light chain protein
NIHSS – National Institute of Health Stroke Scale
NLRP3 – NACHT, LRR, and PYD-domain containing protein 3
NMN – nicotinamide mononucleotide
NR – nicotinamide riboside
NRI – net reclassification index
NT-proBNP – N-terminal fragment B-type natriuretic peptide
OCSP – Oxfordshire community stroke project
OPA1 – mitochondrial dynamin-like GTPase
OR – odds ratio
ORIGIN – outcome reduction with initial glargine intervention
OXA1L – OXA1L mitochondrial inner membrane protein
PACIFIST – Preventing cArdiovascular ComplIcations aFter Ischemic STroke
PCR – polymerase chain reaction
PCSK9 – proprotein convertase subtilin/kexin type 9
PEA – proximity extension assay
Phe-MR – phenome-wide MR
PNPT1 – polyribonucleotide nucleotidyltransferase 1
POLG – DNA polymerase subunit gamma
POLG2 – DNA polymerase subunit gamma-2
PP – posterior probability
PPRC1 – PPARG related coactivator 1

pQTL – protein quantitative trait loci
QMDiab – Qatar metabolomics study on diabetes
qPCR – quantitative polymerase chain reaction
RCT – randomized controlled trial
RRM2B – ribonuclease reductase regulatory TP53 inducible subunit M2B
rt-PA – recombinant tissue plasminogen activator
SAIGE – Scalable and Accurate Implementation of GEneralized mixed mode
SAMHD1 – SAM and HD domain containing deoxynucleoside triphosphate triphosphoydrolase 1
SCALLOP – Systematic and Combined AnaLysis of Olink Proteins
SCARA5 – scavenger receptor class A5
SCO2 – synthesis of cytochrome C oxidase 2
SD – standard deviation
SiGN – Stroke Genetics Network (SiGN)
siRNA – small interfering ribonucleic acid
SNP – single nucleotide polymorphism
SLC25A10 – solute carrier family 25 member 10
STREGA – strengthening the reporting of genetic association studies
STROBE – strengthening the reporting of observational studies
TBRG4 – transforming growth factor beta regulator 4
TFAM – mitochondrial transcription factor A
TNFSF12 – tumor necrosis factor-like weak inducer of apoptosis
TNFRSF12A – tumor necrosis factor receptor superfamily member 12A
TOAST – trial of org 10172 in acute stroke treatment
tRNALeu – transfer ribonucleic acid leucine
TWNK – twinkle mtDNA helicase
TYMP – thymidine phosphorylase
TK2 – thymidine kinase 2
UK – united kingdom
UTR – untranslated region
WES – whole exome sequencing
YFS – young Finns study

## DECLARATION OF ACADEMIC ACHIEVEMENT

FORMAT AND ORGANIZATION OF THESIS

This thesis is prepared in the "sandwich" format as outlined in the School of Graduate Studies' Guide for the Preparation of Theses. It includes a general introduction, an overview of hypotheses and objectives, three independent studies prepared in journal article format, and an overall discussion. The candidate is the first author on all manuscripts. At the time of thesis preparation, Chapter 3 was published, and Chapters 4 and 5 were submitted for review in peer-reviewed journals.

## CONTRIBUTION TO PAPERS WITH MULTIPLE AUTHORSHIP

### Chapter 3 (Study 1)

**Chong M,** Sjaarda J, Pigeyre M, Mohammedi-Shemirani P, Lali R, Gerstein H, Shoamanesh A, Paré G. Novel drug targets for ischemic stroke identified through Mendelian Randomization analysis of the blood proteome. Circulation. 140(10):819-830. (2019)

Author Contributions:

**MC** and GP contributed to the conception and design of the study. **MC**, JS, HG, and GP contributed to data acquisition. **MC**, JS, MP, and PM conducted data analysis. GP facilitated project administration and supervision. **MC** was the principal writer of the manuscript. All authors contributed to the drafting and revision of the final article. All authors approved the final submitted version of the manuscript.

**Chapter 4 (Study 2)**

**Chong M**, Mohammadi-Shemirani P, Nelson W, Perrot N, Morton R, Machipsa T, Lali R, Narula S, Khan M, Khan I, Judge C, Pigeyre M, Akhabir L, O'Donnell M, Paré G. GWAS and ExWAS of blood Mitochondrial DNA copy number identifies 71 loci and highlights a potential causal role in dementia. eLife. Under Review. (2021)

Author Contributions:

**MC** and GP contributed to the conception and design of the study. **MC**, WN, and GP contributed to methodological development. **MC**, PM, NP, GP, SN, IK RL, MK, and TM conducted bioinformatic and/or statistical analyses. GP, MO, and NC contributed to funding or data acquisition. GP facilitated project administration and supervision. **MC** was the principal writer of the manuscript. All authors contributed to the drafting and revision of the final article. All authors approved the final submitted version of the manuscript.

**Chapter 5 (Study 3)**

**Chong M**, Narula S, Morton RW, Judge C, Akhabir L, Cawte N, Pathan N, Lali R, Mohammadi-Shemirani P, Shoamanesh A, O'Donnell M, Yusuf S, Langhorne P, Paré G. Mitochondrial DNA copy number as a marker and mediator of stroke prognosis: Observational and Mendelian Randomization analyses. JAMA Neurology. Submitted. (2021).

Author Contributions:

**MC** and GP contributed to the conception and design of the study. **MC**, WN, and GP contributed to methodological development. **MC**, PL, SN, RM, CJ, LA, NC, NP, RL, PM, AS, MO, SY, and GP contributed to data acquisition, analysis, and/or interpretation of data. GP, MO, and SY obtained funding for this project. CJ, MO, and PL contributed to administrative, technical, or material support. GP, MO, and NC contributed to data acquisition. GP facilitated supervised this project. **MC** was the principal writer of the manuscript. All authors contributed to the drafting and revision of the final article. All authors approved the final submitted version of the manuscript.

# CHAPTER 1:

# INTRODUCTION

# CHAPTER 1: INTRODUCTION

## 1.1 STROKE OVERVIEW

### 1.1.1 STROKE AND ITS IMPACT ON SOCIETY

Stroke, an acute neurological deficit caused by decreased blood supply to the brain, represents the second leading cause of death (11.8% of all deaths) and the third most common cause of disability (4.5% of all life years lost to disability) worldwide[1]. In addition to the tremendous toll on quality-of-life, stroke carries a substantial economic burden with reported costs of $74,353 (CAD) per patient and totals of nearly $2.8 billion annually in Canada alone[2]. Management and treatment are resource intensive as stroke patients carry a significantly higher life-time risk of recurrent stroke and other cardiovascular disease, infection, disability, and dementia[3–6]. A deeper understanding of the molecular risk factors mediating stroke may lead to better risk stratification, prevention, and treatment.

### 1.1.2 STROKE SUBTYPES

Acute neurological deficits may arise from several mechanisms that cause reduced cerebral perfusion. Approximately 70% of all strokes involve vessel blockage and are classified as “ischemic” strokes, and the remaining 30% of strokes are caused by the rupture of cerebral blood vessels leading to intracranial bleeding, also known as “hemorrhagic” stroke[7,8]. The diagnostic workup for determining stroke subtypes is extensive and encompasses a combination of neurological assessment, vascular imaging, neuroimaging (CT or MRI), structural and electrophysiological cardiac testing, and laboratory testing[9]. Most analyses performed in this thesis will focus on ischemic stroke, which will be the focus of the following sections.

Primary stroke types can be further subtyped based on etiology and location. Subtypes of ischemic stroke entail small artery occlusion, large artery atherosclerosis, cardioembolic stroke, stroke of other determined etiology, and strokes of undetermined sources[10]. Small artery occlusion arises from lipohyalinosis or microatheromas within the small penetrating cerebral arteries that directly perfuse brain parenchyma[11]. Lipohyalinosis is a consequence of chronic hypertension and is defined by endothelial dysfunction, local inflammation, and vessel wall thickening which culminate in vascular narrowing[11]. Large artery atherosclerosis refers to the presence of atherosclerotic plaques occluding larger intracranial and/or extracranial arteries or serving as a proximal atherothrombotic embolic source[12]. Cardioembolic stroke is characterized by blood clots originating from cardiac chambers or valves (e.g. left atrium) that subsequently migrate (embolize) to occlude cerebral arteries[13]. Strokes of other determined etiology consist of identifiable albeit rare causes of stroke (e.g. vascular dissection)[10]. Strokes of undetermined source are an active area of investigation but are suspected to be caused by various mechanisms including but not limited to embolism from overlooked cardiac sources (e.g. left ventricular thrombi), paradoxical embolism, and non-occlusive atherosclerosis[14]. In contrast to ischemic stroke subtypes, hemorrhagic stroke subtypes are defined based on whether bleeding occurs in the brain parenchyma (intracerebral hemorrhage) or the space surrounding the brain tissue (subarachnoid hemorrhage).

### 1.1.3 TREATMENT

#### 1.1.3.1 IMPORTANCE OF STROKE SUBTYPING

Accurate diagnosis of subtypes is important because it informs both acute and secondary treatments. In the acute setting for example, recombinant tissue plasminogen activator (rt-PA) is used to degrade clots present in ischemic stroke patients; however, rt-PA also increases the risk of bleeding, and thus is strongly contraindicated for hemorrhagic strokes[15]. For stroke prevention, antithrombotic agents are mainstay therapies since they target a common etiological pathway in thrombosis; however, considerations for the specific class of drug prescribed, timing, and dosage, are made in context of ischemic subtype[16,17]. For example, oral anticoagulation is recommended over antiplatelet therapy in atrial fibrillation patients at high risk of thromboembolism[18].

#### 1.1.3.2 ACUTE STROKE TREATMENT

The adage, “time is brain”, is often used to stress the urgency with which stroke patients should be treated to avoid substantial neuronal loss. It has been estimated that patients with large artery occlusion lose approximately 120 million neurons for every hour that passes by without treatment, which is equivalent to 3.6 years of “brain aging” [19]. As such, the acute phase of ischemic stroke is extremely important to salvage neurons and is centered around timely restoration of blood supply through the disintegration or removal of the blood clot. Thrombolytic therapy entails administration of rt-PA intravenously within 4.5 hours of stroke onset[9]. rt-PA catalyzes the conversion of plasminogen to plasmin, which degrades the fibrin mesh that stabilizes the blood clot[20]. Beyond the 4.5-hour therapeutic window, rt-PA is not effective. However, a subset of strokes with proximal large artery

occlusions affecting the anterior circulation are eligible for mechanical thrombectomy within 24 hours of stroke onset[21]. Mechanical thrombectomy consists of the physical removal of the blood clot from circulation using a stent-retriever device[22]. Major challenges associated with both thrombolysis and thrombectomy include a greater risk for intracranial bleeding and restrictive therapeutic windows[23,24].

#### 1.1.3.3 PRIMARY AND SECONDARY STROKE PREVENTION

In patients with vascular risk factors or who have already suffered a stroke, prevention of future strokes consists of risk factor management and antithrombotic therapy. The importance of controlling stroke risk factors was delineated by the "Importance of Conventional and Emerging Risk Factors of Stroke in Different Regions and Ethnic Groups of the World (INTERSTROKE)"[25,26]. INTERSTROKE found that 10 established risk factors account for approximately 90% of stroke risk globally. These 10 risk factors include hypertension, diabetes mellitus, smoking, alcohol consumption, physical inactivity, dyslipidemia, a diet low in fruits and vegetables, obesity, psychosocial factors, and heart conditions (atrial fibrillation or flutter, myocardial infarction, rheumatic valve disease or prosthetic heart valve). Hypertension is one of the strongest risk factors for stroke overall, and its presence is associated with three and four-fold increased odds of ischemic stroke and intracerebral hemorrhage, respectively[26].

While some risk factors are shared across stroke types, others are more strongly associated with specific subtypes. For example, atrial fibrillation is a common arrhythmia defined by irregular and intermittently stagnant blood flow within the left atria of the heart[27]. Atrial fibrillation potentiates cardiac thrombus formation thereby increasing risk of

cardioembolic stroke by approximately 16-fold[28]. Treatment of atrial fibrillation consists of medical interventions to sustain proper sinus rhythm and use of anticoagulation to prevent future embolic events[29]. Similarly, dyslipidemia is an important risk factor for atherosclerosis, and cholesterol-lowering medication in the form of statins (HMG-COA reductase inhibitors) are employed to treat large artery atherosclerosis[30].

Antithrombotic therapies block the formation of new blood clots, which are composed of (i) an aggregate of platelets known as the "platelet plug" and (ii) a fibrin mesh that stabilizes the platelet plug[31]. The most common antiplatelet therapy is acetylsalicylic acid (aspirin) acts by inhibiting platelet cyclooxygenase (COX1) to impede platelet activation and aggregation[32]. In contrast, anticoagulants block components of the coagulation cascade which constitutes a series of enzymatic cleavages involving protein coagulation factors that ultimately lead to the production of fibrin[18]. For example, a reduced form of vitamin K acts as a cofactor for multiple coagulation factors, and some anticoagulation strategies (e.g. warfarin) work by blocking hepatic vitamin K epoxidase reductase complex[33]. However, inhibition of vitamin K blocks several coagulation factors (II, VII, IX, and X) resulting in impaired hemostasis and greater risk for bleeds[34]. Newer anticoagulants (direct oral anticoagulants) directly target specific coagulation factors, such as factors II (dabigatran) and X (rivaroxaban, apixaban, edoxaban)[18]. Additionally, combined antiplatelet and anticoagulant therapy has recently shown efficacy for stroke prevention in patients with stable atherosclerotic disease[35]. As compared to aspirin monotherapy, the Cardiovascular Outcomes for People using Anticoagulation Strategies

(COMPASS) trial observed superiority for the combination of low-dose rivaroxaban and aspirin for a reduced risk of recurrent ischemic stroke by 67%[36].

## 1.2 STROKE BIOMARKERS

### 1.2.1 RATIONALE FOR STUDY OF STROKE BIOMARKERS

Despite current antithrombotic therapies and risk factor management, the residual risk for recurrent stroke is high (1 to 15% annually)[37–40]. As such, there is a need to uncover novel therapeutic targets. Biological markers (biomarkers) are objective, measurable parameters that can be used for disease prediction, diagnosis, and prognosis[41]. Notably, biomarkers also represent therapeutic targets; for example, many current stroke medications target circulating molecules such as rt-PA (plasmin), aspirin (platelet COX1), and oral anticoagulants (coagulation factors) [18,20,32,42]. In the hours and days following stroke, there are quantifiable changes in thousands of circulating proteins including pro-inflammatory cytokines, chemokines, adhesion molecules, and brain injury markers[43]. Among these circulating biomarkers, there may be a subset that causally mediate stroke pathogenesis and recovery which represent strong candidates for therapeutic targeting. Therefore, elucidating the biomarkers that causally mediate stroke recurrence is likely to reveal novel therapeutic targets for stroke management. In the following sections, we will summarize relevant literature regarding stroke biomarker research.

### 1.2.2 BIOMARKERS FOR STROKE RISK PREDICTION

Stroke risk stratification informs clinical decision making, and biomarkers may help to refine an individual's risk assessment. Circulating C-reactive protein detected through a

high sensitivity assay (hsCRP) is a well-established marker of systemic inflammation and is associated with greater incidence of both ischemic and hemorrhagic stroke[44–46]. Similarly, higher circulating levels of the pro-inflammatory cytokine, Interleukin-6 (IL6), predict greater incidence of ischemic stroke though this increased risk is entirely mediated by differences in risk factor profiles[46,47]. LP(a) is a highly atherogenic, pro-inflammatory, and pro-thrombotic circulating lipoprotein. A large cohort study including more than 60,000 Danish individuals showed a strong association between elevated LP(a) and stroke incidence (HR=1.60; 95% CI, 1.24-2.05) after adjustment for known risk factors including LDL cholesterol[48]. An even larger study of 283,540 British individuals confirmed this relationship and showed modest improvement in discrimination of atherosclerotic cardiovascular disease events by incorporating Lp(a) (c-index 0.640 vs. 0.642)[49].

A novel set of risk scores that integrate circulating biomarkers has shown promise for stroke and bleeding risk prediction in atrial fibrillation patients, known as the "age, biomarker and clinical history" (ABC) scores[50,51]. The ABC-stroke score incorporates two circulating cardiac biomarkers associated with incident ischemic stroke: N-terminal fragment B-type natriuretic peptide (NT-proBNP) and cardiac troponin T[51]. NT-proBNP is secreted by cardiomyocytes in response to myocyte stretch and is a marker of heart failure and an independent predictor of cardioembolic stroke[52]. Cardiac troponin T is released into circulation acutely after myocardial injury and elevated levels are used to diagnose myocardial infarction[53]. The ABC-stroke risk score better discriminates stroke risk when compared to the established $CHA_2DS_2$-VASc risk score which does not incorporate biomarkers (c-index 0.66 vs. 0.58) and performs well in both anticoagulated and non-

coagulated patients with atrial fibrillation [51,54]. Complementary to the ABC-stroke score is the ABC-bleeding score for the prediction of major bleeding, a key adverse side-effect of anticoagulation therapy[50,55]. The biomarkers incorporated into the ABC-bleeding score are growth differentiation factor 15 (GDF15), cardiac troponin T, and hemoglobin. Circulating GDF15 is often upregulated in response to stressors, and levels are prognostic for major bleeding and death[50,56]. The ABC-bleeding score outperforms the conventional HAS-BLED score for risk discrimination of major bleeds (c-index 0.68 vs. 0.61)[50].

### 1.2.3 BIOMARKERS FOR STROKE PROGNOSTICATION

Several circulating proteins have been reported to be associated with stroke prognosis including those implicated in inflammation (IL6, hsCRP, S100A8/A9), vascular remodelling (matrix metallopeptidase 9 [MMP9]), and brain injury (neurofilament light chain [NFl]). The modified Rankin Scale (mRS) is an ordinal metric used to capture post-stroke functional outcome and ranges from “no symptoms at all” (mRS 0) to “death” (mRS 6) (Table 1.1)[57]. Circulating IL6 is predictive of both stroke incidence and prognosis, and higher levels at hospital admission portends worse functional outcome (mRS 3-6 vs. 0-2) at 3-months in both ischemic and hemorrhagic stroke patients[58]. Serum levels of another inflammatory mediator, complement C3, correlate with poor functional outcome 3-months post-ischemic stroke and improve reclassification of individuals with poor (mRS 3-6) vs. good (mRS 0-2) functional outcomes (NRI=0.09)[59]. The prognostic utility of a multi-biomarker risk score was investigated in 3,575 participants from the China Antihypertensive Trial in Acute Ischemic Stroke (CATIS) trial[60]. Specifically, the addition of hsCRP, GDF15, MMP9, and S100A8/A9 to conventional risk factors (age, sex, fasting

plasma glucose, lipids, systolic blood pressure, time from onset to hospitalisation, cigarette smoking status, alcohol drinking status, baseline stroke severity [NIHSS], ischemic stroke subtype, and antihypertensive medication post-admission) significantly improved reclassification of poor functional outcome at 3-months post-stroke (mRS 3-6 vs. 0-2; NRI=0.33). MMP9 plays an important role in remodelling of the extracellular matrix following stroke, and S100A8/A9 is abundantly present in neutrophils and monocytes which are the first responders of the immune defence to inflammation[61,62]. Finally, NFl is exclusively expressed in the axonal cytoskeleton of neurons and is only released into circulation after brain injury, and thus has been dubbed as the "neurologist's troponin"[63]. NFl shows prognostic utility for multiple neurodegenerative disorders and stroke[63]. Elevated plasma NFl predicts poor functional outcome (mRS 3-6) at 3-months for patients with cardioembolic stroke, strokes of undetermined source, subarachnoid hemorrhage, and intracerebral hemorrhage[64].

**Table 1.1.** The modified Rankin Scale (mRS) as a measure of disability and dependence[57].

| Score | Definition |
|---|---|
| 0 | No symptoms at all |
| 1 | No significant disability: despite symptoms, able to carry out all usual duties and activities |
| 2 | Slight disability: unable to perform all previous activities but able to look after own affairs without assistance |
| 3 | Moderate disability: requiring some help but able to walk without assistance |
| 4 | Moderately severe disability: unable to walk without assistance and unable to attend to own bodily needs without assistance |
| 5 | Severe disability: bedridden, incontinent and requiring constant nursing care and attention |
| 6 | Death |

### 1.2.4 NEW HORIZONS FOR STROKE BIOMARKER RESEARCH

#### 1.2.4.1 MITOCHONDRIAL DNA COPY NUMBER (MTDNA-CN)

Mitochondria are semi-autonomous organelles that have important roles in stroke pathogenesis and recovery[65,66]. In the presence of ischemia, hypoxic neurons transition from oxidative to anaerobic phosphorylation, which eventually fails to meet the energy demands of the cell. As a result, ATP-dependent ion channels which normally maintain the electrochemical gradient between intracellular and extracellular compartments lose their function. An influx of sodium and calcium ions ensues, triggering mitochondrial swelling and the release of pro-apoptotic factors from the mitochondrial membrane (cytochrome C) which then induces neuronal cell death.

A seminal proteomics study by Garcia-Berrrocoso *et al* (2018) further highlights the critical role of mitochondrial pathways in the response to cerebral ischemia[67]. Laser microdissection was used to isolate the neurovasculature of deceased ischemic stroke patients, and proteomic changes in infarcted brain tissue were compared to contralateral tissue. Mitochondrial dysfunction and oxidative phosphorylation constituted the top pathways that were differently expressed between infarcted and non-infarcted neurons. Additionally, a novel neuroprotective mechanism involving the intercellular transfer of mitochondria has been described in a rodent stroke model[68]. During stroke, astrocytes transfer their mitochondria to oxygen-deprived neurons, eliciting pro-survival signals. As a result, this is accompanied by several beneficial effects including smaller infarcts (less severe strokes) and better functional outcomes. A separate protective mechanism has also been described in endothelial progenitor cells, which release functionally viable

mitochondria into circulation[69]. These extracellular mitochondria are then taken up by damaged brain endothelium which partially restores the integrity of the blood brain barrier.

Despite evidence supporting a strong role for mitochondria in stroke pathogenesis and recovery, mitochondrial biomarkers are seldom studied in human stroke patients. An emerging and inexpensive marker of mitochondrial activity is mitochondrial DNA copy number (mtDNA-CN)[70,71]. In the blood, most mitochondria are located in white blood cells, platelets, and various extracellular sources. Hitherto, investigations have measured mtDNA-CN predominantly from buffy coat samples, in which case, mtDNA-CN denotes the ratio between mitochondrial and autosomal DNA copies within white blood cells (and platelets if platelet-depletion is not performed). Sometimes viewed as a simple marker of mitochondrial abundance, lower white blood cell mtDNA-CN levels are thought to represent states of general mitochondrial dysfunction, oxidative stress, and inflammation[72].

While all three phenomena are relevant to stroke, evidence from genetic, experimental, and epidemiological studies also support a direct role of mtDNA-CN in stroke. First, patients with rare genetic disorders characterized by very low mtDNA-CN, known as "mtDNA depletion syndromes", experience leukoencephalopathy and stroke-like episodes[73,74]. mtDNA depletion syndromes are caused by genetic defects in enzymes involved in nucleotide metabolism, mtDNA replication, and mtDNA repair[75]. Second, experimental studies suggest that mtDNA-CN is a mediator of ischemia reperfusion injury and stroke recovery[76–78]. Indeed, rats with experimentally induced ischemic stroke (middle cerebral artery occlusion) experience a precipitous drop in mtDNA-CN levels, and pre-stroke mtDNA-CN levels can be recovered by injection of a cleavage-resistant form of a

key mtDNA regulator, Optic Atrophy 1 (OPA1)[77]. The recovery in mtDNA-CN levels via OPA1 also attenuated stroke severity and improved functional outcomes in rats. Third, epidemiological analyses demonstrate that low leukocyte mtDNA-CN is associated with stroke risk in humans. A retrospective case-control study found that stroke cases have lower leukocyte mtDNA-CN than controls and that this coincides with higher levels of oxidative stress markers[79]. A large meta-analysis of four prospective cohorts totalling 20,163 participants including 1,583 incident stroke events (mean follow-up of 13.5 years) showed that low mtDNA-CN levels at baseline predicted higher risk for incident stroke[80]. In other disease settings (heart disease, peripheral artery disease, chronic kidney disease, and sudden cardiac death), low leukocyte mtDNA-CN is associated with incident risks of secondary hospitalization, infection, and mortality[80–83]. Hitherto, mtDNA-CN has not been extensively studied as a prognostic marker nor as therapeutic target for stroke in humans[84].

#### 1.2.4.2 ADVANCES IN MULTIPLEX PROTEIN DETECTION ENABLE PROTEOME-WIDE DISCOVERY

Most biomarker studies have evaluated only a select subset of proteins with prior evidence of (i) mediating stroke pathogenesis in animal models (e.g. MMP9), (ii) being associated with related neurological diseases in epidemiological studies (e.g. NFl), or (iii) capturing established risk factor pathways (e.g. hsCRP). However, this "candidate" biomarker approach is inherently limited in discovery potential and overlooks thousands of unique circulating proteins with less well-characterized functions. Surveying the entire circulating proteome is met by several challenges including (i) a wide dynamic range of concentration spanning 12 orders of magnitude (femtomolar to milligram), (ii) the vastness

of the circulating proteome encompassing ~10,000 unique proteins, and (iii) the non-specificity of traditional immunoassays due to antibody cross-reactivity. Innovative technological breakthroughs have increased the number of measurable proteins from up to hundreds of proteins using traditional multiplex immunoassays to thousands. For large-scale epidemiological investigations, two high-throughput proteomics technologies have been widely adopted, SOMAlogic's Slow Off-rate Modified Aptamers (SOMAmers) technology and OLINK's Proximity Extension Assay (PEA) technology[85,86].

SOMAmer technology currently offers the broadest coverage of the proteome enabling detection of 7,000 unique circulating proteins (https://somalogic.com/somascan-discovery/). Aptamers are short oligonucleotides with high affinity towards a specific protein epitope that are generated through an iterative amplification process: a pool of random oligonucleotides encounters a target protein, and the aptamer with the highest affinity to a specific protein epitope is amplified[86]. After multiple rounds of aptamer selection and amplification, chemical modifications to stabilize the aptamer are added in the form of chemical modifications that behave like amino acid side chains[86]. The resulting molecule is a SOMAmer. SOMAmers are conjugated to fluorophores, which renders them amenable to multiplex protein quantification on a fluorescent microarray. Key limitations of SOMAmer technology include non-specificity (cross-reactive binding to non-target human proteins or even non-human proteins given that SOMAmers are not "natural" human antibodies) and variable detection depending on SNP variation (35% of SOMAmers display altered binding affinity in the presence of SNPs)[87].

The OLINK PEA technology transduces protein quantification into a DNA quantification problem (Figure 1.1)[85]. Each protein is targeted by a pair of antibodies, and each antibody is labelled by a single-stranded DNA probe whose sequence is complementary to the cognate antibody's single-stranded DNA probe. When antibody pairs bind the target protein, their complementary single-stranded DNA labels come into close proximity, forming a double-stranded DNA complex that primes an extension reaction. The resulting full-length DNA barcode is amplified by PCR and can then be quantified via quantitative PCR or next-generation sequencing (Figure 1.1). While more limited in proteomic coverage (up to ~1500 proteins) as compared to the SOMAmer detection method, the main advantage of PEA technology is its specificity (https://www.olink.com/products/olink-explore/). High specificity is maintained in the presence of cross-reactive antibody binding since the DNA labels of mismatching antibodies do not form double-stranded DNA complexes and thus cross-reactive events do not contribute to quantifiable signal.

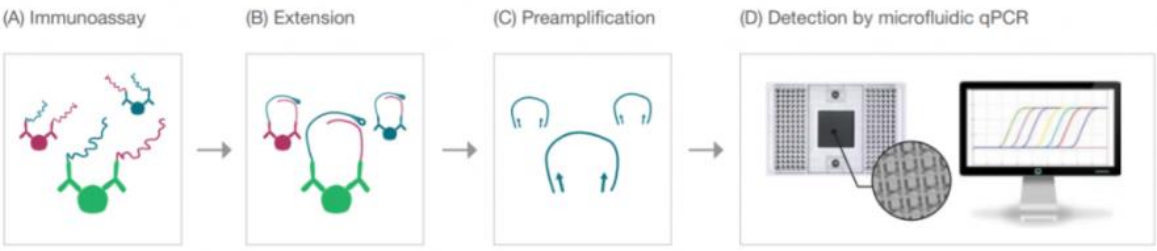



**Figure 1.1.** OLINK PEA technology overview. (Screenshot from https://www.olink.com/data-you-can-trust/technology/)

## 1.3 INTEGRATING GENETICS & BIOMARKERS FOR TARGET DISCOVERY

### 1.3.1 ENHANCING TARGET PRIORITIZATION WITH GENETICS

Recent advances in multiplex protein assay technology have equipped researchers with the ability to interrogate thousands of plasma proteins in large epidemiological studies, and while this greatly expands the scope for therapeutic discovery, a central challenge remains in being able to pinpoint true drug targets among disease-associated biomarkers. The steep cost of developing new medications ($1.2 billion CAD per drug) and high attrition rates (> 50% drug candidates fail for lack of efficacy in RCTs) emphasize the utility of effective drug prioritization strategies since not all disease-associated biomarkers cause disease[88,89]. For example, hsCRP is an acute phase reactant protein whose circulating levels increase in response to activation of the NLRP3 inflammasome-IL1-IL6 axis[46]. Essentially, CRP is a downstream product of this inflammation pathway, thus serving as a clinically useful surrogate for inflammation levels. However, interventions reducing hsCRP levels *per se* are not expected to mitigate inflammation. Genetic studies provide evidence against a causal role for hsCRP in vascular inflammation, and conversely, support for a causal role for an upstream regulator of CRP, IL6[90–93]. A single nucleotide polymorphism (SNP; rs7553007) proximal to the *CRP* gene is associated with serum CRP concentration (i.e. cis pQTL) but not risk of heart disease or stroke[46]. In contrast, SNPs within the IL6 pathway are associated with increased inflammation as well as increased risk of heart disease and stroke[91,92]. Indeed, a seminal randomized controlled trial showed that canakinumab, a monoclonal antibody that blocks the upstream activator of IL6, IL1β, reduces recurrent risk of cardiovascular events and death in myocardial infarction patients

with elevated hsCRP[94]. RCTs are currently planned to test pharmacologic agents (Anakinra and colchicine) to target the NRLP3-IL1 inflammasome axis for secondary cardiovascular disease prevention in ischemic stroke (Preventing cArdiovascular ComplIcations aFter Ischemic STroke [PACIFIST]; Colchicine for prevention of Vascular Inflammation in Non-CardioEmbolic Stroke [CONVINCE]) and intracerebral hemorrhage (Colchicine for the prevention of vascular events after an acute IntraCerebral Hemorrhage [CoVasc-ICH]) patients (private communication with Dr. Askhan Shoamanesh). In summary, genetic associations can be used to clarify causality between disease-biomarker associations, thereby providing a means to prioritize causal mediators as potential therapeutic targets.

Therapeutic potential is hinted at in contexts where the genetic determinants of biomarkers and stroke risk overlap[95]. The intuition is that if a biomarker causally affects stroke risk, then genetic variants that influence biomarker levels should also have a corresponding and proportional effect on stroke risk. By using genetic evidence to infer causality of biomarker-stroke relationships, the subset of associated biomarkers causally mediating risk of stroke likely represents stronger drug candidates. Indeed, a review by Nelson *et al.* (2015) showed that pharmacological compounds with support from genetic association studies are more than twice as likely to reach market approval as compared to those without genetic support[96].

### 1.3.2 MENDELIAN RANDOMIZATION (MR) ANALYSIS

Mendelian Randomization (MR) analysis is a statistical genetics framework for causal inference that uses genetic variants as "instruments" to approximate the unconfounded effect of an exposure (e.g. biomarker) on an outcome (e.g. stroke) [97,98,99].

Similar to how RCTs randomize interventions, MR leverages the natural randomization of genetic alleles that occurs during meiosis (i.e. Mendel's second law of independent assortment)[100] (Figure 1.2). While epidemiological associations may be susceptible to confounders and reverse causation, the randomization of exposure-associated alleles ensures that confounders are balanced between allele carriers and non-carriers. Furthermore, because genetic variants are inherited, the flow of cause-and-effect is unidirectional from genetic variant to change in biomarker level to disease consequence thus conferring some protection against reverse causation. In relation to stroke, MR has been mainly used to clarify whether a causal relationship exists between well-established clinical risk factors and risk of stroke. Such studies demonstrate causal roles for blood pressure and total stroke, adiposity and ischemic stroke, atrial fibrillation and ischemic stroke, type 2 diabetes and small artery occlusion, and LDL cholesterol and large artery atherosclerosis[101–103]. Candidate blood biomarkers have also been assessed through the MR framework with positive findings for lipoprotein(a) and matrix-metalloproteinase-12 (MMP12) but not hsCRP, cystatin C, or YKL-40[90,91,104–107].

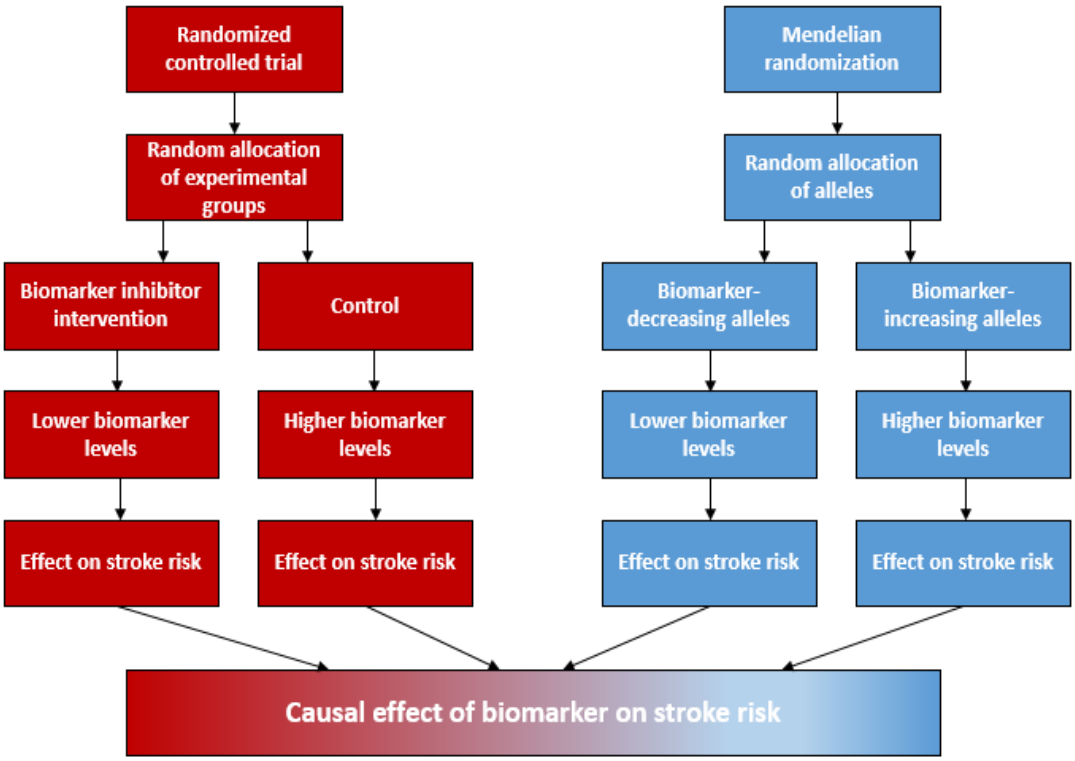
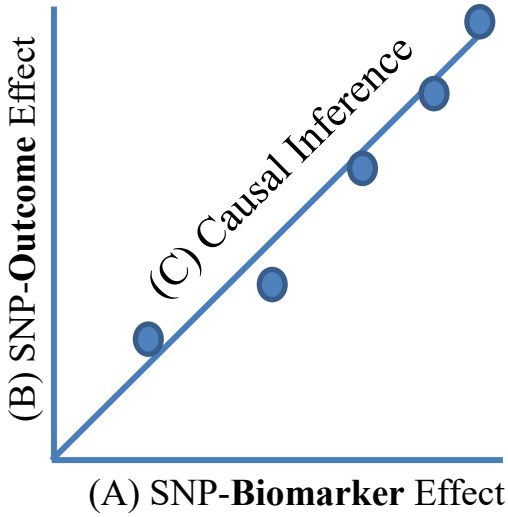



**Figure 1.2** A comparison of randomized controlled trials with the Mendelian Randomization study design.

Causality is supported when the effects of multiple, independent genetic variants on biomarker levels are directionally consistent and proportional to their effects on stroke risk (Figure 1.3). The inputs necessary to conduct MR analysis are two sets of genetic effects: (A) the effect of genetic variants on biomarker levels and (B) the effect of genetic variants on stroke risk. Understanding the relationship between these effects permits triangulation of the causal effect of biomarker levels on stroke risk (C). Genetics effects can be derived through a genetic study design known as a genome-wide association study (GWAS) allowing for comprehensive interrogation of common variants dispersed across the genome. Notably, only summary-level data, in the form of effect estimates and their

standard errors, is required to perform MR analysis[98,108]. This statistical methodological advance in the MR field has led to an explosion in the number of studies of this kind




conducted in the past 5 years.

**Figure 1.3.** Visualization of the underlying premise for MR as a causal inference method. Each point represents an independent genetic variant.

### 1.3.3 GENOME-WIDE ASSOCIATION STUDIES (GWAS)

GWAS are systematic and agnostic surveys of common variant associations that allows for the identification of specific genetic determinants of traits and more specifically, an approximation of the effect of each genetic variant on the trait in question[109]. Following a similar trajectory as biomarker investigations, genetic studies began with "candidate" gene approaches but were supplanted by GWAS once multiplex SNP detection technology was developed, namely, the SNP microarray[110]. Microarrays contain millions of microwells that each harbour an individual assay for a unique SNP site, and as such, genotypes for hundreds of thousands to several millions of SNPs can be measured simultaneously in multiple samples[110]. Statistical genetics advances in the form of imputation techniques

leverage the correlation structure (linkage disequilibrium) between variants in large aggregates of sequencing data (imputation referenc) to boost the number of detectable genetic variants (genomic coverage) to approximately 10 million common SNPs[111–113].

#### 1.3.3.1 GWAS FOR CIRCULATING BIOMARKERS

Recent large-scale studies combining multiplex genomics and proteomics technologies highlight a major role for genetics in the regulation of circulating protein levels. A seminal study by Sun *et al.* (2018) investigated genetic determinants for 3,622 plasma proteins using SOMAmer technology in 3,301 healthy blood donors from the INTERVAL study[105]. This study identified 1,927 associations for 1,104 unique proteins which represented a four-fold increase in the number of protein quantitative trait loci (pQTL) known at that time. Approximately 30% of pQTLs were located within or nearby the genes encoding for the circulating protein, also known as cis-pQTLs. Considering the vastness of the genome, this reinforces a strong role for proximal genetic regulation. Since the publication of this study, international research collaborations have been convened to exhaustively characterize the genetic determinants of circulating proteins. Most notably, the Systematic and Combined AnaLysis of Olink Proteins (SCALLOP) consortium has recently published GWAS of PEA-detected proteins associated with cardiovascular diseases including up to 30,000 study participants[114].

#### 1.3.3.2 GWAS FOR STROKE

Finding robust genetic associations for stroke through GWAS was initially challenging. The very first GWAS for ischemic stroke was conducted in 2009 by Ikram *et al.* in 19,602 participants (1164 ischemic strokes) and revealed *NINJ2* as a putative stroke

locus, but this association has never been replicated[115]. This failure was attributed to overlooking the vast phenotypic heterogeneity of stroke and relatively small sample sizes. As Hacke *et al*. expressed in a GWAS commentary, many studies have fallen victim to the fact that stroke represents an umbrella term for etiologically distinct stroke subtypes[116]. Essentially, stroke is a syndrome encompassing many distinct clinical entities with a shared predisposition for thromboembolism.

Since the first stroke GWAS in 2009, several collaborative initiatives have been convened to conduct large GWAS meta-analyses including METASTROKE (N=90,648) and the NINDS Stroke Genetics Network (SiGN) (N=435,001) for ischemic stroke subtypes[117,118]. A seminal GWAS of stroke (N=521,612) was published in March 2018 by the MEGASTROKE consortium[119]. This study represented a major leap forward in our understanding of the genetic architecture of ischemic stroke subtypes. Prior to this study, only 10 reliable stroke loci were known. MEGASTROKE not only replicated these loci but also uncovered 22 new loci for a total of 32 loci. Of the 32 loci, 13 (41%) demonstrated subtype specificity and the remaining loci were either associated with multiple ischemic subtypes independently or broader stroke types (ischemic stroke or any stroke). A follow-up analysis by Traylor *et al*. (2019) further investigated subtype specificity for a subset of 16 MEGASTROKE loci and concluded that the pattern of subtype association for loci was highly heterogeneous, though seven (44%) of the 16 loci were found to influence both hemorrhagic and ischemic stroke risk suggesting a shared etiology between the major stroke types[120]. Furthermore, many MEGASTROKE loci were also associated with well-established vascular risk factors including venous thromboembolism, lipids, coronary

artery disease, blood pressure, carotid plaque, and atrial fibrillation (Figure 1.4). Intriguingly, 11 loci were not known to influence classic risk factors suggesting novel mechanisms for stroke pathophysiology and are now being investigated more closely. For example, Traylor *et al.* (2020) further investigated the *PDE3A* variant in an independent study and observed that mutation carriers had impaired flow-mediated dilatation, a marker of endothelial reactivity[121].

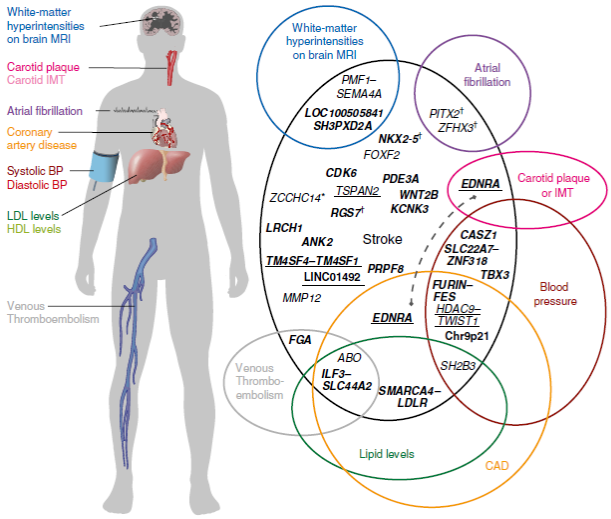



**Figure 1.4.** A pictorial representation of the 32 stroke loci identified by the MEGASTROKE study and the overlap with vascular risk factors and traits[119].

Post-MEGASTROKE, GWAS have focused on underrepresented stroke subtypes (e.g. small vessel disease), radiological markers (e.g. perivascular dilated spaces), and stroke outcomes (e.g. 3-month mRS)[122–125]. Indeed, stroke outcome genetics is one of the major research priorities for future stroke genetics research according to the International Stroke Genetics Consortium (ISGC)[126]. To this end, the Genetics of Ischaemic Stroke Functional Outcome (GISCOME) performed the first international GWAS of 3-month

mRS and identified the first GWAS locus for ischemic stroke outcome, an intronic variant in the *LOC105372028* locus (rs1842681)[125]. A second phase GWAS meta-analysis is under way to uncover additional genetic determinants of post-stroke outcome and therapeutic response (private communications).

### 1.3.4 EMERGING STROKE THERAPIES WITH MR SUPPORT

GWAS is the backbone of MR analysis as they are used to identify genetic variants associated with human traits, which can then be used as genetic instruments to evaluate causality with outcomes. Advances in GWAS for both circulating biomarkers and stroke provide an exciting opportunity for drug target discovery at an unprecedented scale[105,114,119,127–130]. Indeed, several new stroke therapies targeting coagulation and dyslipidemia have been recently supported by MR evidence[42]. For example, antisense oligonucleotide molecules blocking coagulation factor XI have been developed and are undergoing testing in RCTs for secondary stroke prevention and safety profiling in atrial fibrillation patients (NCT04304508, NCT03582462, NCT04218266). Complementary to these ongoing RCTs is a MR analysis by Georgi *et al.* (2020) showing that genetically lower FXI levels is associated with reduced risk of venous thrombosis and cardioembolic stroke[131]. Proprotein convertase subtilin/kexin type 9 (PCSK9) is an enzyme that facilitates the endocytosis and degradation of membrane-bound low-density lipoprotein receptor (LDLR) which clears LDL cholesterol from circulation[132–137]. Several monoclonal antibodies have been developed to inhibit PCSK9 production as a novel means of cholesterol lowering, and two RCTs have demonstrated that when PCSK9 inhibition is added to statin therapy, risk of recurrent stroke risk is further reduced by 25%[138,139]. MR

also supports a causal role for genetically reduced PCSK9 in ischemic stroke risk reduction[99]. Intriguingly, RCTs also observe that PCSK9 inhibition is associated with a reduction in another independent atherogenic mediator, Lipoprotein(a) (Lp(a)), suggesting potentially pleiotropic effects of PCSK9 inhibition beyond cholesterol-lowering[140,141]. A hepatocyte-directed antisense oligonucleotide therapy for Lp(a) (AKCEA-APO(a)-$L_{Rx}$) reduces circulating Lp(a) by approximately 80% as shown in a recent phase II trial, but whether this translates to clinical benefit remains to be determined in ongoing phase III trials. One of the earliest stroke MR studies evaluated LP(a) due to its high heritability and the fact that a single cis-pQTL (rs10455872) accounted for a significant proportion (~30%) of the variance in plasma concentrations[48,104]. MR analyses substantiate a strong role for genetically lower Lp(a) in protecting against atherosclerotic disease including heart disease, aortic valve stenosis, and large artery atherosclerosis[142]. Altogether, the fact that MR supports a causal role for several emerging therapeutic targets is both reassuring for ongoing phase III trials and corroborates the utility of MR as a drug prioritization tool.

## 1.4 REFERENCES


1. Feigin, V. L., Norrving, B. & Mensah, G. A. Global Burden of Stroke. *Circ. Res.* **120**, 439–448 (2017).
2. Mittmann, N. *et al.* Impact of disability status on ischemic stroke costs in Canada in the first year. *Can. J. Neurol. Sci.* **39**, 793–800 (2012).
3. Mohan, K. M. *et al.* Risk and cumulative risk of stroke recurrence: a systematic review and meta-analysis. *Stroke* **42**, 1489–94 (2011).

4. Gunnoo, T. *et al.* Quantifying the risk of heart disease following acute ischaemic stroke: a meta-analysis of over 50,000 participants. *BMJ Open* **6**, e009535 (2016).
5. Westendorp, W. F., Nederkoorn, P. J., Vermeij, J. D., Dijkgraaf, M. G. & van de Beek, D. Post-stroke infection: A systematic review and meta-analysis. *BMC Neurol.* **11**, 110 (2011).
6. Leys, D., Hénon, H., Mackowiak-Cordoliani, M. A. & Pasquier, F. Poststroke dementia. *Lancet Neurol.* **4**, 752–759 (2005).
7. Krishnamurthi, R. V *et al.* Stroke Prevalence, Mortality and Disability-Adjusted Life Years in Adults Aged 20-64 Years in 1990-2013: Data from the Global Burden of Disease 2013 Study. *Neuroepidemiology* **45**, 190–202 (2015).
8. Thrift, A. G., Dewey, H. M., Macdonell, R. A. L., McNeil, J. J. & Donnan, G. A. Incidence of the major stroke subtypes initial findings from the North East Melbourne Stroke Incidence Study (NEMESIS). *Stroke* **32**, 1732–1738 (2001).
9. Powers, W. J. *et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke a guideline for healthcare professionals from the American Heart Association/American Stroke A. Stroke* vol. 50 (2019).
10. Adams, H. P. *et al.* Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* **24**, 35–41 (1993).
11. Regenhardt, R. W. *et al.* Pathophysiology of Lacunar Stroke: History's Mysteries and Modern Interpretations. *J. Stroke Cerebrovasc. Dis.* **28**, 2079–2097 (2019).

12. Cole, J. W. Large artery occlusive disease. *Continuum (N. Y).* **23**, 133–157 (2017).

13. Kamel, H. & Healey, J. S. Cardioembolic Stroke. *Circ. Res.* **120**, 514–526 (2017).

14. Hart, R. G. *et al.* Embolic strokes of undetermined source: The case for a new clinical construct. *Lancet Neurol.* **13**, 429–438 (2014).

15. Fugate, J. E. & Rabinstein, A. A. Absolute and Relative Contraindications to IV rt-PA for Acute Ischemic Stroke. *The Neurohospitalist* **5**, 110–121 (2015).

16. Yadav, J. S.; Wholey, M. H.; Kuntz, R. E., Fayad, Pierre; Katzen, B. T.; Mishkel, G. J.; Bajwa, T. K.; hitlow, P.; Strickman, N. E.; Jaff, M. R.; Popma, J.J.; Snead, D. B.; Cutlip, D. E.; Firth, B. G.; Ouriel, K. . Rivaroxaban versus Warfarin in Nonvalvular Atrial Fibrillation. *N. Engl. J. Med.* **351**, 1493–1501 (2004).

17. Kwok, C. S. *et al.* Efficacy of Antiplatelet Therapy in Secondary Prevention Following Lacunar Stroke: Pooled Analysis of Randomized Trials. *Stroke* **46**, 1014–1023 (2015).

18. Chen, A., Stecker, E. & A Warden, B. Direct Oral Anticoagulant Use: A Practical Guide to Common Clinical Challenges. *J. Am. Heart Assoc.* **9**, e017559 (2020).

19. Saver, J. L. Time is brain - Quantified. *Stroke* **37**, 263–266 (2006).

20. Collen, D. Molecular mechanism of action of newer thrombolytic agents. *J. Am. Coll. Cardiol.* **10**, 11B-15B (1987).

21. Jadhav, A. P. *et al.* Eligibility for Endovascular Trial Enrollment in the 6- to 24-Hour Time Window: Analysis of a Single Comprehensive Stroke Center. *Stroke* **49**, 1015–1017 (2018).

22. Smith, W. S. *et al.* Mechanical thrombectomy for acute ischemic stroke: Final

results of the multi MERCI trial. *Stroke* **39**, 1205–1212 (2008).

23. Miller, D. J., Simpson, J. R., Silver, B. & Silver, B. Safety of Thrombolysis in Acute Ischemic Stroke: A Review of Complications, Risk Factors, and Newer Technologies. *The Neurohospitalist* **1**, 138–147 (2011).

24. Meinel, T. R. *et al.* Endovascular stroke treatment and risk of intracranial hemorrhage in anticoagulated patients. *Stroke* 892–898 (2020) doi:10.1161/STROKEAHA.119.026606.

25. O'Donnell, M. J. *et al.* Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* **376**, 112–23 (2010).

26. O'Donnell, M. J. *et al.* Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. *Lancet* **388**, 761–775 (2016).

27. Michaud, G. F. & Stevenson, W. Atrial fibrillation. *N Engl J Med* **384**, 353–361 (2021).

28. Griñán, K. *et al.* Cardioembolic Stroke: Risk Factors, Clinical Features, and Early Outcome in 956 Consecutive Patients. *Rev. Invest. Clin.* **73**, 023–030 (2020).

29. January, C. T. *et al.* 2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart R. *Circulation* **140**, e125–e151 (2019).

30. Safouris, A. *et al.* Statin pretreatment and microembolic signals in large artery atherosclerosis a systematic review and meta-analysis. *Stroke* **49**, 1992–1995 (2018).

31. Hankey, G. J. Antithrombotic Therapy for Stroke Prevention: What's New? *Circulation* **139**, 1131–1133 (2019).

32. Hackam, D. G. & Spence, J. D. Antiplatelet therapy in ischemic stroke and transient ischemic attack: An overview of major trials and meta-analyses. *Stroke* **50**, 773–778 (2019).

33. Wu, S. *et al.* Warfarin and Vitamin K epoxide reductase: a molecular accounting for observed inhibition. *Blood* **132**, 647–657 (2018).

34. Mijares, M. E., Nagy, E., Guerrero, B. & Arocha-Piñango, C. L. [Vitamin K: biochemistry, function, and deficiency. Review]. *Invest. Clin.* **39**, 213–29 (1998).

35. Eikelboom, J. W. *et al.* Rivaroxaban with or without Aspirin in Stable Cardiovascular Disease. *N. Engl. J. Med.* NEJMoa1709118 (2017) doi:10.1056/NEJMoa1709118.

36. Sharma, M. *et al.* Stroke Outcomes in the COMPASS Trial. *Circulation* **139**, 1134–1145 (2019).

37. Sps, T. & Group, S. Blood-pressure targets in patients with recent lacunar stroke: The SPS3 randomised trial. *Lancet* **382**, 507–515 (2013).

38. Dwyer, D. J., Camacho, D. M., Kohanski, M. A., Callura, J. M. & Collins, J. J. Stenting versus Aggressive Medical Therapy for Intracranial Arterial Stenosis. **46**, 561–572 (2013).

39. Brott, T. G., Hobson, R. W. & Howard, G. Stenting versus Endarterectomy for Treatment of Carotid-Artery Stenosis. *J. Vasc. Surg.* **52**, 799 (2010).

40. Patel, M. R. *et al.* Rivaroxaban versus Warfarin in Nonvalvular Atrial Fibrillation. *N. Engl. J. Med.* **365**, 883–891 (2011).

41. Califf, R. M. Biomarker definitions and their applications. *Exp. Biol. Med.* **243**, 213–221 (2018).

42. Katsanos, A. H. & Hart, R. G. New Horizons in Pharmacologic Therapy for Secondary Stroke Prevention. *JAMA Neurol.* **77**, 1308–1317 (2020).

43. Montaner, J. *et al.* Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke. *Nat. Rev. Neurol.* **16**, 247–264 (2020).

44. Kaptoge, S. *et al.* C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: An individual participant meta-analysis. *Lancet* **375**, 132–140 (2010).

45. Karim, M. A. *et al.* Systemic inflammation is associated with incident stroke and heart disease in East Asians. *Sci. Rep.* **10**, 1–11 (2020).

46. Ridker, P. M. From CRP to IL-6 to IL-1: Moving Upstream To Identify Novel Targets for Atheroprotection. *HHS Public Access* **118**, 145–156 (2016).

47. Jenny, N. S. *et al.* Inflammatory cytokines and ischemic stroke risk: The REGARDS cohort. *Neurology* **92**, E2375–E2384 (2019).

48. Langsted, A., Nordestgaard, B. G. & Kamstrup, P. R. Elevated Lipoprotein(a) and Risk of Ischemic Stroke. *J. Am. Coll. Cardiol.* **74**, 54–66 (2019).

49. Trinder, M., Uddin, M. M., Finneran, P., Aragam, K. G. & Natarajan, P. Clinical

Utility of Lipoprotein(a) and LPA Genetic Risk Score in Risk Prediction of Incident Atherosclerotic Cardiovascular Disease. *JAMA Cardiol.* **6**, 287–295 (2021).

50. Hijazi, Z. *et al.* The novel biomarker-based ABC (age, biomarkers, clinical history)-bleeding risk score for patients with atrial fibrillation: a derivation and validation study. *Lancet* **387**, 2302–2311 (2016).

51. Hijazi, Z. *et al.* The ABC (age, biomarkers, clinical history) stroke risk score: A biomarker-based risk score for predicting stroke in atrial fibrillation. *Eur. Heart J.* **37**, 1582–1590 (2016).

52. Kim, H. N. & Januzzi, J. L. Natriuretic peptide testing in heart failure. *Circulation* **123**, 2015–2019 (2011).

53. Chapman, A. R., Bularga, A. & Mills, N. L. High-Sensitivity Cardiac Troponin Can Be An Ally in the Fight Against COVID-19. *Circulation* 1–8 (2020) doi:10.1161/circulationaha.120.047008.

54. Benz, A. P. *et al.* Biomarker-Based Risk Prediction With the ABC-AF Scores in Patients With Atrial Fibrillation Not Receiving Oral Anticoagulation. *Circulation* **143**, 1863–1873 (2021).

55. Oyama, K. *et al.* Serial assessment of biomarkers and the risk of stroke or systemic embolism and bleeding in patients with atrial fibrillation in the ENGAGE AF-TIMI 48 trial. *Eur. Heart J.* 1–9 (2021) doi:10.1093/eurheartj/ehab141.

56. Hijazi, Z. *et al.* A biomarker-based risk score to predict death in patients with atrial fibrillation: The ABC (age, biomarkers, clinical history) death risk score. *Eur.*

*Heart J.* **39**, 477–485 (2018).

57. Banks, J. L. & Marotta, C. A. Outcomes Validity and Reliability of the Modified Rankin Scale : Implications for Stroke Clinical Trials A Literature Review and Synthesis. (2007) doi:10.1161/01.STR.0000258355.23810.c6.

58. Mechtouff, L. *et al.* Association of Interleukin-6 levels and futile reperfusion after mechanical thrombectomy. *Neurology* 10.1212/WNL.0000000000011268 (2020) doi:10.1212/WNL.0000000000011268.

59. Yang, P. *et al.* Increased Serum Complement C3 Levels Are Associated with Adverse Clinical Outcomes after Ischemic Stroke. *Stroke* 868–877 (2021) doi:10.1161/STROKEAHA.120.031715.

60. Guo, D. *et al.* Prognostic Metrics Associated with Inflammation and Atherosclerosis Signaling Evaluate the Burden of Adverse Clinical Outcomes in Ischemic Stroke Patients. *Clin. Chem.* **66**, 1434–1443 (2020).

61. Montaner, J. *et al.* Matrix metalloproteinase expression is related to hemorrhagic transformation after cardioembolic stroke. *Stroke* **32**, 2762–2767 (2001).

62. Wang, S. *et al.* S100A8/A9 in inflammation. *Front. Immunol.* **9**, (2018).

63. Thebault, S., Booth, R. A. & Freedman, M. S. Blood neurofilament light chain: The neurologist's troponin? *Biomedicines* **8**, 1–11 (2020).

64. Gendron, T. F. *et al.* Plasma neurofilament light predicts mortality in patients with stroke. *Sci. Transl. Med.* **12**, 19–26 (2020).

65. Liu, F., Lu, J., Manaenko, A., Tang, J. & Hu, Q. Mitochondria in ischemic stroke: New insight and implications. *Aging Dis.* **9**, 924–937 (2018).

66. Chen, W., Guo, C., Feng, H. & Chen, Y. Mitochondria: Novel Mechanisms and Therapeutic Targets for Secondary Brain Injury After Intracerebral Hemorrhage. *Front. Aging Neurosci.* **12**, 1–10 (2021).

67. García-Berrocoso, T. *et al.* Single cell immuno-laser microdissection coupled to label-free proteomics to reveal the proteotypes of human brain cells after ischemia. *Mol. Cell. Proteomics* **17**, 175–189 (2018).

68. Hayakawa, K. *et al.* Transfer of mitochondria from astrocytes to neurons after stroke. *Nature* **535**, 551–555 (2016).

69. Hayakawa, K. *et al.* Protective effects of endothelial progenitor cell-derived extracellular mitochondria in brain endothelium. *Stem Cells* **36**, 1404–1410 (2019).

70. Robin, E. D. & Wong, R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J. Cell. Physiol.* **136**, 507–513 (1988).

71. Malik, A. N. & Czajka, A. Is mitochondrial DNA content a potential biomarker of mitochondrial dysfunction? *Mitochondrion* **13**, 481–492 (2013).

72. Variations, H., Duan, M. & Tu, J. Recent Advances in Detecting Mitochondrial DNA. 1–18 (2018) doi:10.3390/molecules23020323.

73. Paramasivam, A. *et al.* Homozygous R627W mutations in POLG cause mitochondrial DNA depletion leading to encephalopathy, seizures and stroke-like episodes. *Mitochondrion* **48**, 78–83 (2019).

74. Bonora, E. *et al.* Biallelic variants in LIG3 cause a novel mitochondrial neurogastrointestinal encephalomyopathy. *Brain* 1–17 (2021) doi:10.1093/brain/awab056.

75. Gorman, G. S. *et al.* Mitochondrial diseases. *Nat. Rev. Dis. Prim.* **2**, (2016).

76. Varanita, T. *et al.* The Opa1-Dependent Mitochondrial Cristae Remodeling Pathway Controls Atrophic , Apoptotic , Article The Opa1-Dependent Mitochondrial Cristae Remodeling Pathway Controls Atrophic , Apoptotic , and Ischemic Tissue Damage. 834–844 (2015) doi:10.1016/j.cmet.2015.05.007.

77. Lai, Y. *et al.* Restoration of L-OPA1 alleviates acute ischemic stroke injury in rats via inhibiting neuronal apoptosis and preserving mitochondrial function. *Redox Biol.* **34**, 101503 (2020).

78. Zhao, M. *et al.* Mitochondrial ROS promote mitochondrial dysfunction and inflammation in ischemic acute kidney injury by disrupting TFAM-mediated mtDNA maintenance. *Theranostics* **11**, (2021).

79. Lien, L. *et al.* Significant Association Between Low Mitochondrial DNA Content in Peripheral Blood Leukocytes and Ischemic Stroke. *J. Am. Heart Assoc.* (2017) doi:10.1161/JAHA.117.006157.

80. Ashar, F. N. *et al.* Association of Mitochondrial DNA Copy Number With Cardiovascular Disease. **21205**, 1247–1255 (2017).

81. Fazzini, F. *et al.* Mitochondrial DNA copy number is associated with mortality and infections in a large cohort of patients with chronic kidney disease. **8**, 480–488 (2019).

82. Zhang, Y. *et al.* Association between mitochondrial DNA copy number and sudden cardiac death : findings from the Atherosclerosis Risk in Communities study ( ARIC ). 3443–3448 (2017) doi:10.1093/eurheartj/ehx354.

83. Song, L. *et al.* mtDNA Copy Number Contributes to All-Cause Mortality of Lacunar Infarct in a Chinese Prospective Stroke Population. (2019).

84. Song, L. *et al.* mtDNA Copy Number Contributes to All-Cause Mortality of Lacunar Infarct in a Chinese Prospective Stroke Population. *J. Cardiovasc. Transl. Res.* **13**, 783–789 (2020).

85. Assarsson, E. *et al.* Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* **9**, (2014).

86. Davies, D. R. *et al.* Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19971–19976 (2012).

87. Joshi, A. & Mayr, M. In Aptamers They Trust: Caveats of the SOMAscan Biomarker Discovery Platform from SomaLogic. *Circulation* **138**, 2482–2485 (2018).

88. Wouters, O. J., McKee, M. & Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA - J. Am. Med. Assoc.* **323**, 844–853 (2020).

89. Arrowsmith, J. & Miller, P. Trial Watch: Phase II and Phase III attrition rates 2011-2012. *Nat. Rev. Drug Discov.* **12**, 569 (2013).

90. Prins, B. P. *et al.* Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. *PLoS Med.* **13**, 1–29 (2016).

91. Freitag, D. *et al.* Cardiometabolic effects of genetic upregulation of the interleukin

1 receptor antagonist: A Mendelian randomisation analysis. *Lancet Diabetes Endocrinol.* **3**, 243–253 (2015).

92. Georgakis, M. K. *et al.* Genetically Downregulated Interleukin-6 Signaling Is Associated with a Favorable Cardiometabolic Profile: A Phenome-Wide Association Study. *Circulation* 1177–1180 (2021) doi:10.1161/CIRCULATIONAHA.120.052604.

93. Thériault, S. *et al.* Genetic Association Analyses Highlight IL6, ALPL, and NAV1 As 3 New Susceptibility Genes Underlying Calcific Aortic Valve Stenosis. *Circ. Genomic Precis. Med.* **12**, 431–441 (2019).

94. Ridker, P. M. *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N. Engl. J. Med.* **377**, 1119–1131 (2017).

95. Gill, D. *et al.* Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Res.* **6**, (2021).

96. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).

97. Smith, G. D. & Hemani, G. Mendelian randomization: Geneticanchorsfor causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, 89–98 (2014).

98. Hemani, G. *et al.* The MR-base platform supports systematic causal inference across the human phenome. *Elife* **7**, 1–29 (2018).

99. Rao, A. S. *et al.* Large-Scale Phenome-Wide Association Study of PCSK9 Variants Demonstrates Protection Against Ischemic Stroke. *Circ. Genomic Precis. Med.* **11**, e002162 (2018).

100. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **27**, R195–R208 (2018).

101. Hopewell, J. C. & Clarke, R. Emerging Risk Factors for Stroke: What Have We Learned from Mendelian Randomization Studies? *Stroke* **47**, 1673–1678 (2016).

102. Liu, J., Rutten-Jacobs, L., Liu, M., Markus, H. S. & Traylor, M. Causal Impact of Type 2 Diabetes Mellitus on Cerebral Small Vessel Disease. *Stroke* **49**, 1325–1331 (2018).

103. Hindy, G. *et al.* Role of blood lipids in the development of ischemic stroke and its subtypes: A mendelian randomization study. *Stroke* **49**, 820–827 (2018).

104. Helgadottir, A. *et al.* Apolipoprotein(a) genetic sequence variants associated with systemic atherosclerosis and coronary atherosclerotic burden but not with venous thromboembolism. *J. Am. Coll. Cardiol.* **60**, 722–729 (2012).

105. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).

106. Russi, A. E. & Brown, M. A. Cystatin C and Cardiovascular Disease: A Mendelian Randomization Study. **165**, 255–269 (2016).

107. Kjaergaard, A. D., Johansen, J. S., Bojesen, S. E. & Nordestgaard, B. G. Elevated plasma YKL-40, lipids and lipoproteins, and ischemic vascular disease in the general population. *Stroke* **46**, 329–335 (2015).

108. Bowden, J. & Holmes, M. V. Meta-analysis and Mendelian randomization: A review. *Res. Synth. Methods* **10**, 486–496 (2019).

109. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019).

110. Laframboise, T. Single nucleotide polymorphism arrays : a decade of biological , computational and technological advances. **37**, 4181–4193 (2009).

111. Browning, B. L. & Browning, S. R. Genotype Imputation with Millions of Reference Samples. *Am. J. Hum. Genet.* **98**, 116–126 (2016).

112. Durbin, R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics* **30**, 1266–72 (2014).

113. Zachary, A. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Biorxiv* 1–46 (2019).

114. Folkersen, L. *et al.* Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* **2**, 1135–1148 (2020).

115. Ikram, M. A. *et al.* Genomewide Association Studies of Stroke. *N. Engl. J. Med.* **360**, 1718–1728 (2009).

116. Hacke, W. & Grond-Ginsbach, C. Commentary on a GWAS: HDAC9 and the risk for ischaemic stroke. *BMC Med.* **10**, 70 (2012).

117. Traylor, M. *et al.* Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet. Neurol.* **11**, 951–62 (2012).

118. Rosand, J. *et al.* Loci associated with ischaemic stroke and its subtypes (SiGN): A genome-wide association study. *Lancet Neurol.* **15**, 174–184 (2016).

119. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects

identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).

120. Traylor, M. *et al.* Subtype Specificity of Genetic Loci Associated With Stroke in 16 664 Cases and 32 792 Controls. *Circ. Genomic Precis. Med.* **12**, 307–314 (2019).

121. Traylor, M. *et al.* Influence of genetic variation in PDE3A on endothelial function and stroke. *Hypertension* 365–371 (2020) doi:10.1161/HYPERTENSIONAHA.119.13513.

122. Chung, J. *et al.* Genome-wide association study of cerebral small vessel disease reveals established and novel loci. *Brain* **142**, 3176–3189 (2019).

123. Traylor, M. *et al.* Articles Genetic basis of lacunar stroke : a pooled analysis of individual patient data and genome-wide association studies. *Lancet Neurol.* **4422**, 1–11 (2021).

124. Persyn, E. *et al.* Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nat. Commun.* **11**, 1–12 (2020).

125. Soderholm, M. *et al.* Genome-wide association meta-analysis of functional outcome after ischemic stroke. *Neurology* **0**, 1271–1283 (2019).

126. Woo, D. *et al.* Top research priorities for stroke genetics. *Lancet Neurol.* **17**, 663–665 (2018).

127. Yao, C. *et al.* Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat. Commun.* **9**, (2018).

128. Nath, A. P. *et al.* Multivariate Genome-wide Association Analysis of a Cytokine

Network Reveals Variants with Widespread Immune, Haematological, and Cardiometabolic Pleiotropy. *Am. J. Hum. Genet.* **105**, 1076–1090 (2019).

129. Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science (80-. ).* **361**, 769–773 (2018).

130. Malik, R. *et al.* Genome-wide meta-analysis identifies 3 novel loci associated with stroke. *Ann. Neurol.* **84**, 934–939 (2018).

131. Georgi, B. *et al.* Leveraging Human Genetics to Estimate Clinical Risk Reductions Achievable by Inhibiting Factor XI. *Stroke* **50**, 3004–3012 (2019).

132. Abifadel, M. *et al.* Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.* **34**, 154–156 (2003).

133. Austin, M. A., Hutter, C. M., Zimmern, R. L. & Humphries, S. E. Genetic causes of monogenic heterozygous familial hypercholesterolemia: a HuGE prevalence review. *Am. J. Epidemiol.* **160**, 407–20 (2004).

134. Abul-Husn, N. S. *et al.* Genetic identification of familial hypercholesterolemia within a single U.S. Health care system. *Science (80-. ).* **354**, (2016).

135. Rousselet, E. *et al.* PCSK9 reduces the protein levels of the LDL receptor in mouse brain during development and after ischemic stroke. *J. Lipid Res.* **52**, 1383–1391 (2011).

136. Poirier, S. *et al.* The proprotein convertase PCSK9 induces the degradation of low density lipoprotein receptor (LDLR) and its closest family members VLDLR and ApoER2. *J. Biol. Chem.* **283**, 2363–2372 (2008).

137. Seidah, N. G., Awan, Z., Chrétien, M. & Mbikay, M. PCSK9: A key modulator of

cardiovascular health. *Circ. Res.* **114**, 1022–1036 (2014).

138. Marston, N. A. *et al.* Predicting Benefit from Evolocumab Therapy in Patients with Atherosclerotic Disease Using a Genetic Risk Score. *Circulation* 616–623 (2020) doi:10.1161/CIRCULATIONAHA.119.043805.

139. GG, S. *et al.* Alirocumab and Cardiovascular Outcomes after Acute Coronary Syndrome. *N Engl J Med* **379**, 2097–2107 (2018).

140. O'Donoghue, M. L. *et al.* Lipoprotein(a), PCSK9 inhibition, and cardiovascular risk insights from the FOURIER trial. *Circulation* **139**, 1483–1492 (2019).

141. Szarek, M. *et al.* Lipoprotein(a) lowering by alirocumab reduces the total burden of cardiovascular events independent of low-density lipoprotein cholesterol lowering:ODYSSEY OUTCOMES trial. *Eur. Heart J.* **41**, 4245–4255 (2020).

142. Larsson, S. C. *et al.* Lipoprotein(a) in Alzheimer, Atherosclerotic, Cerebrovascular, Thrombotic, and Valvular Disease: Mendelian Randomization Investigation. *Circulation* **141**, 1826–1828 (2020).

# CHAPTER 2:

# GENERAL HYPOTHESIS, OBJECTIVE, RATIONALE, AND APPROACH

# CHAPTER 2: HYPOTHESIS, OBJECTIVE, RATIONALE, & APPROACH

## 2.1 GENERAL HYPOTHESIS

We hypothesize that a subset of blood biomarkers causally mediate stroke risk and prognosis.

## 2.2 GENERAL OBJECTIVE

The overall objective of this PhD thesis is to identify molecular determinants of stroke risk and post-stroke outcomes to generate novel therapeutic targets.

## 2.3 RATIONALE AND APPROACH

A decade ago, neuroprotective agents were heralded as the incumbent class of stroke medications following antithrombotic therapies; however, compelling animal model findings failed to translate to clinical benefit in human patients[1]. This cautionary tale emphasizes the importance of effective drug target prioritization tools that combine evidence from multiple modalities not reliant on a single source. One such promising framework that integrates multiple 'Omics modalities is Mendelian Randomization (MR) analysis[2]. MR analysis has been successfully applied to drug target discovery and validation for numerous diseases via the triangulation of genomic, biomarker, and outcome information[3]. Accordingly, we will apply MR analysis to identify putative targets for stroke treatment in two ways: (i) systematic identification of novel targets (*a priori* discovery) and (ii) assessment of a single targeted hypothesis (*priori* evaluation).

No study has used MR agnostically to identify novel drug targets for ischemic stroke. The recent emergence of large-scale GWAS for thousands of circulating proteins in combination with GWAS of thousands of stroke patients enables such an investigation for

the first time. Accordingly, we will (i) systematically screen the circulating proteome for novel drug targets, (ii) forecast effects of target manipulation on key safety phenotypes (i.e. intracranial bleeding), and (iii) comprehensively elucidate side-effect profiles (Chapter 3).

Mitochondrial dysfunction has long been known to be a sequala and mediator of post-stroke brain injury[4]. However, only recently has an accessible blood marker of mitochondrial activity (leukocyte mtDNA-CN) emerged for study in human participants[5]. Findings from animal models indicate that circulating mtDNA-CN levels acutely drop following stroke, and that rescuing mtDNA-CN levels attenuates stroke severity and improves post-stroke functional outcomes[6]. In contrast, the role of mtDNA-CN in human stroke patients as both a marker and causal determinant of stroke prognosis has not been extensively investigated. The latter query has not yet been addressed mainly because the genetic determinants of mtDNA-CN remain elusive. To this end, we will (i) develop a novel method for array-based mtDNA-CN estimation to enable convenient estimation of mtDNA-CN from biobank-scale genomic datasets and (ii) apply this method to conduct a large GWAS in the UKBiobank study to find genetic variants associated with mtDNA-CN levels (Chapter 4). Finally, we will (i) characterize the epidemiological association between leukocyte mtDNA-CN measured within one week of stroke onset and post-stroke outcomes in patients from the INTERSTROKE study, and (ii) perform MR analysis using independent datasets (UKBiobank from Chapter 4 and GISCOME) to assess whether there is evidence supporting a causal effect of mtDNA-CN on post-stroke outcomes (Chapter 5).

## 2.4 REFERENCES


1. Chamorro, Á., Dirnagl, U., Urra, X. & Planas, A. M. Neuroprotection in acute stroke: Targeting excitotoxicity, oxidative and nitrosative stress, and inflammation. *Lancet Neurol.* **15**, 869–881 (2016).
2. Hemani, G. *et al.* The MR-base platform supports systematic causal inference across the human phenome. *Elife* **7**, 1–29 (2018).
3. Gill, D. *et al.* Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Res.* **6**, (2021).
4. Liu, F., Lu, J., Manaenko, A., Tang, J. & Hu, Q. Mitochondria in ischemic stroke: New insight and implications. *Aging Dis.* **9**, 924–937 (2018).
5. Fazzini, F. *et al.* Plasmid-normalized quantification of relative mitochondrial DNA copy number. 1–11 (2018) doi:10.1038/s41598-018-33684-5.
6. Lai, Y. *et al.* Restoration of L-OPA1 alleviates acute ischemic stroke injury in rats via inhibiting neuronal apoptosis and preserving mitochondrial function. *Redox Biol.* **34**, 101503 (2020).

# CHAPTER 3:

## Novel drug targets for ischemic stroke identified through Mendelian Randomization Analysis of the Blood Proteome

Published in *Circulation.* 140:819-839 (2019)

Circulation

ORIGINAL RESEARCH ARTICLE

# Novel Drug Targets for Ischemic Stroke Identified Through Mendelian Randomization Analysis of the Blood Proteome

Editorial, see p 831

Michael Chong, MSc
Jennifer Sjaarda, PhD
Marie Pigeyre, MD, PhD
Pedrum Mohammadi-Shemirani, BSc
Ricky Lali, MSc
Ashkan Shoamanesh, MD
Hertzel Chaim Gerstein, MD, MSc
Guillaume Paré, MD, MSc

**BACKGROUND:** Novel, effective, and safe drugs are warranted for treatment of ischemic stroke. Circulating protein biomarkers with causal genetic evidence represent promising drug targets, but no systematic screen of the proteome has been performed.

**METHODS:** First, using Mendelian randomization (MR) analyses, we assessed 653 circulating proteins as possible causal mediators for 3 different subtypes of ischemic stroke: large artery atherosclerosis, cardioembolic stroke, and small artery occlusion. Second, we used MR to assess whether identified biomarkers also affect risk for intracranial bleeding, specifically intracerebral and subarachnoid hemorrhages. Third, we expanded this analysis to 679 diseases to test a broad spectrum of side effects associated with hypothetical therapeutic agents for ischemic stroke that target the identified biomarkers. For all MR analyses, summary-level data from genome-wide association studies (GWAS) were used to ascertain genetic effects on circulating biomarker levels versus disease risk. Biomarker effects were derived by meta-analysis of 5 GWAS (N≤20 509). Disease effects were derived from large GWAS analyses, including MEGASTROKE (N≤322 150) and UK Biobank (N≤408 961) studies.

**RESULTS:** Several biomarkers emerged as causal mediators for ischemic stroke. Causal mediators for cardioembolic stroke included histo-blood group ABO system transferase, coagulation factor XI, scavenger receptor class A5 (SCARA5), and tumor necrosis factor–like weak inducer of apoptosis (TNFSF12). Causal mediators for large artery atherosclerosis included ABO, cluster of differentiation 40, apolipoprotein(a), and matrix metalloproteinase-12. SCARA5 (odds ratio [OR]=0.78; 95% CI, 0.70–0.88; $P$=1.46×10$^{-5}$) and TNFSF12 (OR=0.86; 95% CI, 0.81–0.91; $P$=7.69×10$^{-7}$) represent novel protective mediators of cardioembolic stroke. TNFSF12 also increased the risk of subarachnoid (OR=1.53; 95% CI, 1.31–1.78; $P$=3.32×10$^{-8}$) and intracerebral (OR=1.34; 95% CI, 1.14–1.58; $P$=4.05×10$^{-4}$) hemorrhages, whereas SCARA5 decreased the risk of subarachnoid hemorrhage (OR=0.61; 95% CI, 0.47–0.81; $P$=5.20×10$^{-4}$). Multiple side effects beyond stroke were identified for 6 of 7 biomarkers, most (75%) of which were beneficial. No adverse side effects were found for coagulation factor XI, apolipoprotein(a), and SCARA5.

**CONCLUSIONS:** Through a systematic MR screen of the circulating proteome, causal roles for 5 established and 2 novel biomarkers for ischemic stroke were identified. Side-effect profiles were characterized to help inform drug target prioritization. In particular, SCARA5 represents a promising target for treatment of cardioembolic stroke, with no predicted adverse side effects.

**Key Words:** biomarkers ■ genetics ■ intracranial hemorrhages ■ Mendelian randomization analysis ■ proteomics ■ stroke

Sources of Funding, see page 828



https://www.ahajournals.org/journal/circ

ORIGINAL RESEARCH ARTICLE

## Clinical Perspective

### What Is New?

- Among 653 proteins, 7 were causal mediators of ischemic stroke including 2 established targets, apolipoprotein(a) and coagulation factor XI, and 2 novel mediators of cardioembolic stroke: scavenger receptor class A5 (SCARA5) and tumor necrosis factor weak inducer of apoptosis.
- Targeting SCARA5 was predicted to also protect against subarachnoid hemorrhage with no evidence of adverse side effects.
- Some biomarkers mediated risk of multiple nonstroke disorders.

### What Are the Clinical Implications?

- Findings provide confirmatory evidence for pursuing clinical trials of coagulation factor XI and apolipoprotein(a).
- SCARA5 represents a new therapeutic target.
- Integrating genomic, proteomic, and phenomic data through Mendelian randomization facilitates discovery of drug targets and their side effects.

Stroke is a leading cause of death and disability worldwide, and its global burden continues to grow.[1] Randomized controlled trials (RCTs) reveal high stroke recurrence rates among ischemic stroke patients ranging from 1.0% to 14.9% per year, depending on the underlying subtype (ie, large artery atherosclerosis, cardioembolic stroke, or small artery occlusion).[2–4] As such, novel therapies for treatment of ischemic stroke are warranted. Epidemiological studies have identified several circulating proteins that correlate with stroke risk,[5–7] thus representing potential therapeutic targets. However, high attrition rates associated with drug development prompt the need for further investigation of stroke-associated biomarkers before clinical testing.[8] Nelson et al[8] demonstrated that a protein drug target whose link with disease is supported by genetic association is twice as likely to reach market approval.[8] Furthermore, recent technological advances in high-throughput protein quantification have enabled genome-wide association studies (GWAS) to uncover genetic determinants for thousands of blood proteins simultaneously.[9–13] Accordingly, we sought to discover new and effective drug targets for ischemic stroke by integrating genetic and proteomic data through Mendelian randomization (MR) analysis.

MR is a statistical genetics framework used to assess causality between an exposure (ie, biomarker) and an outcome (ie, ischemic stroke).[14] Similar to how RCTs randomly allocate an intervention to test its causal effect on an outcome, MR represents a "natural" RCT that leverages the random allocation of exposure-influencing genetic alleles.[15] Previously, MR has been applied in a hypothesis-driven manner to assess causality of select biomarkers on stroke risk including C–C motif chemokine ligand 2 (CCL2), Chitinase-3–like protein 1 (CHI3L1), C-reactive protein (CRP), cystatin C (CST3), apo lipoprotein(a) (LPA), matrix metalloproteinase-12 (MMP12), and proprotein convertase subtilisin/kexin type 9 (PCSK9).[5–7,12,16–19] However, there has been no systematic scan of the human proteome for novel causal mediators of stroke. Beyond drug target prioritization, MR can also be applied to predict target-mediated side effects to reveal unanticipated adverse effects and opportunities for drug repurposing.[15,19]

In this study, we used MR to (1) systematically screen 653 circulating proteins to identify novel mediators of ischemic stroke subtypes (large artery atherosclerosis, cardioembolic stroke, and small artery occlusion); (2) examine the relationship between identified biomarkers and risk of intracranial bleeding; and (3) predict target-mediated side effects through phenome-wide analysis of 679 disease traits.

## METHODS

### Data Disclosure Statement

The authors declare that supporting data are available within the article, its online supplementary files, and referenced public data sets.

### Deriving Genetic Determinants of Circulating Biomarker Levels

The premise underlying MR is that causality is supported if genetic variants that influence exposure levels (ie, biomarker) also influence the outcome (ie, ischemic stroke) in a proportional and directionally concordant manner. Accordingly, we first identified a subset of genetic variants that were associated with circulating biomarker levels. To accomplish this, we combined 5 different biomarker GWAS analyses for which genome-wide summary statistics were publicly available (YFS/FINRISK [Cardiovascular Risk in Young Finns Study/FINRISK]; IMPROVE [The Multicentre, Longitudinal Carotid Intima-Media Thickness and IMT-Progression as Predictors of Vascular Events in a High-Risk European Population]; KORA F4/QMDiab [Cooperative Health Research in the Region of Augsburg F4/Qatar Metabolomics Study on Diabetes]; and INTERVAL) or that we had access to in-house (ORIGIN [Outcome Reduction with Initial Glargine Intervention]).[9–13] The 5 study samples consisted of predominantly European individuals in whom biomarker testing was conducted in blood samples using various high-multiplex protein assays. Circulating proteins were measured through bead-based immunoassays in YFS/FINRISK and ORIGIN, modified antibodies conjugated to oligonucleotides in IMPROVE, and slow off-rate modified aptamers in KORA F4/QMDiab and INTERVAL. Specifically, YFS/FINRISK analyzed 41 cytokines detected via the Bio-Rad Bio-Plex Pro human cytokine assay; ORIGIN analyzed 227 biomarkers via the Myriad RBM human Explorer multianalyte profile assay; IMPROVE analyzed 83 cardiovascular disease–related proteins via the O-link ProSeek cardiovascular disease array I; KORA

FR/QMDiab analyzed 1124 proteins via the SOMAScan assay (version 3.2); and INTERVAL analyzed 2994 proteins via the SOMAScan assay (extended panel).[9–13] For further details, see Table I in the online-only Data Supplement. To maintain the validity of genetic variants for MR analyses, blood biomarkers possessing any of the following attributes were excluded:

1. Biomarker is encoded by a gene that produces multiple proteins
2. Biomarker is encoded by multiple genes
3. Biomarker is encoded by a nonautosomal gene (eg, X-chromosome encoded)
4. Biomarker is not encoded by a gene (eg, biomarker is a steroid hormone)
5. Biomarker is known to exhibit substantial structural variation (eg, immunoglobulins).

After these exclusions, 3090 unique protein biomarkers remained, of which 899 (29.1%) were assayed by multiple studies. Meta-analysis was subsequently performed using a fixed-effect inverse-variance weighted model in METAL.[20] For each biomarker, common single-nucleotide polymorphisms (SNPs; minor allele frequency > 0.01 in 1000Genomes Europeans) with a combined *P* value below 0.01 were retained. Furthermore, SNPs located beyond 200 kilobases of the encoding gene transcript of the protein biomarker were removed to maximize the likelihood that SNP-biomarker associations were directly mediated through the biomarker itself, thereby mitigating the potential for pleiotropy. This resultant set of biomarker-associated SNPs were used as "genetic instruments" to estimate genetically determined biomarker levels in all subsequent MR analyses. Genetic instruments for all identified biomarkers are available in Table II in the online-only Data Supplement.

## Systematic MR Screening for Causal Mediators of Ischemic Stroke Subtypes

The relationship between biomarker-associated SNPs (identified in the previous GWAS meta-analysis) and ischemic stroke subtypes was assessed using summary-level data from the MEGASTROKE study.[21] Specifically, effect size and standard error estimates were extracted from European-only MEGASTROKE analyses for large artery atherosclerosis ($N_{case}$=4373; $N_{control}$= 200 618), cardioembolic stroke ($N_{case}$=7193; $N_{control}$=314 957), and small artery occlusion ($N_{case}$= 5386; $N_{control}$=249 172). Because genetic instruments for circulating biomarker levels were derived from Europeans (because of the lack of available non-European biomarker GWAS), European-only MEGASTROKE analyses were used in lieu of transethnic analyses. Essentially, ethnic composition of exposure and outcome data sets was harmonized to satisfy the presumption that genetic instruments approximated similar biomarker effects in both exposure and outcome data sets. Ischemic stroke subtyping was based on the TOAST (Trial of Org 10172 in Acute Stroke Treatment) criteria.[4] Biomarker-associated SNPs that were not present in MEGASTROKE were excluded. For the remaining SNPs with available effect size estimates for both biomarker levels and risk of stroke, linkage disequilibrium-pruning ($r^2$<0.10 in 1000Genomes Europeans) was performed to generate an independent set of genetic variants.[22] Analyses in which genetic variants collectively explained less than 0.5% of variance in biomarker levels were excluded to ensure that only robust associations would withstand multiple hypothesis testing.

## Statistical Analysis

The MR–robust adjusted profile scoring method was chosen for MR testing because of its resilience to violations of key MR assumptions, such as horizontal pleiotropy.[23] Furthermore, MR–robust adjusted profile scoring enables inclusion of many weak instruments by accounting for the precision of SNP-exposure associations in addition to SNP-outcome associations. Odds ratios were expressed per SD increase in genetically determined circulating biomarker levels. A Bonferroni-corrected *P* value threshold accounting for both the number of biomarkers and outcomes analyzed was implemented (Figure 1; $P=0.05/(653\times3)=2.55\times10^{-5}$).

Standard sensitivity analyses were employed to assess the validity of MR findings, including verification with orthogonal MR methods (inverse variance weighted, MR–Egger, and median weighted MR), as well as diagnostic tests for pleiotropy (MR–Egger intercept), directionality (Steiger), and SNP outliers (leave-one-out analysis). All MR analyses were completed using the "MRBase for TwoSample MR" package (version 0.4.09).[24] The plots were generated using various R (version 3.4.1) packages including ggplot2, MRBase, and PheWAS. It should be noted that the same statistical analysis framework, incorporating MR–robust adjusted profile scoring and the aforementioned sensitivity analyses, was employed in all subsequent MR analyses as well.

## Targeted MR Analysis of Hemorrhagic Stroke Subtypes

Next, given that ischemic and hemorrhagic stroke subtypes share common clinical and genetic risk factors with evidence for both concordant and antagonistic effects,[21,25] using MR, we sought to characterize the causal effects of ischemic stroke biomarkers on 2 hemorrhagic subtypes: subarachnoid and intracerebral hemorrhage (Figure 1). SNP-biomarker effects were derived from the same GWAS meta-analysis as in the primary analysis for ischemic stroke subtypes (N=20 509). SNP-outcome effects were derived by performing association testing of genetic instruments within British participants from the large, prospective UK Biobank study who consented for genetic analysis (application 15255). Quality control of the genetic data set was performed as previously mentioned by Sjaarda et al,[13] except that related individuals were retained in the present analysis. Association testing was conducted using the SAIGE (Scalable and Accurate Implementation of GEneralized mixed model; version 0.29) method developed by Zhou et al[26] to account for unbalanced case-control ratios. Subarachnoid hemorrhage (N=1064) and intracerebral hemorrhage (N=845) cases were classified using algorithmic definitions based on a combination of hospital admission, self-reported, and death register data (UK Biobank Field IDs: 42010 and 42012).[27] Controls (N=398 736) for this analysis consisted of stroke-free individuals (UK Biobank Field ID: 42006). Covariates and model parameters followed those in Zhou et al.[26] The results were expressed as odds ratios per SD increase in genetically determined circulating biomarker levels. As a secondary analysis, intracerebral hemorrhage subtypes (lobar and nonlobar) were also investigated using GWAS summary statistics from Woo et al[28] ($N_{lobar}$=664; $N_{nonlobar}$=881; $N_{control}$=1481).

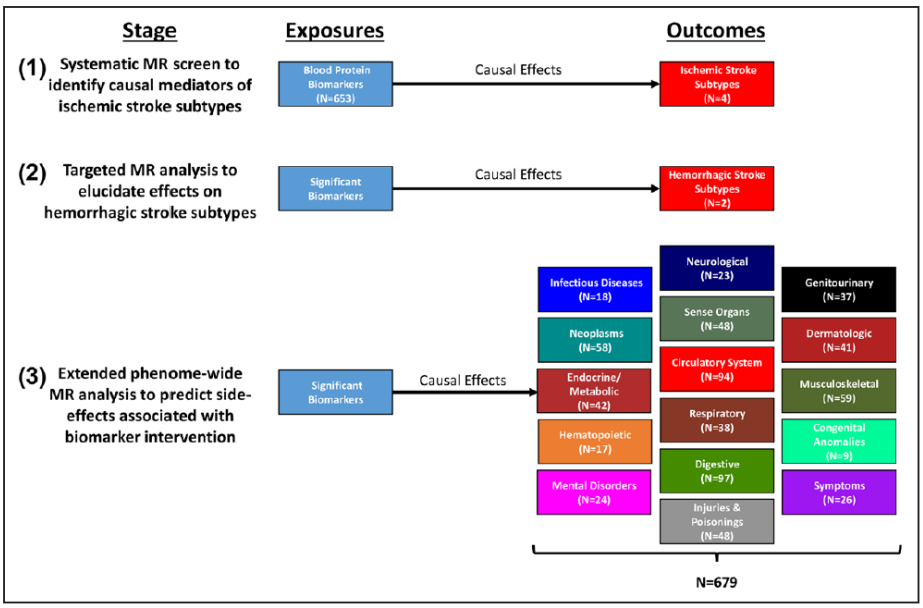



**Figure 1. Overview of Mendelian randomization (MR) analyses.**
The study consists of a 3-stage design that employs MR at all stages. First, we evaluated causal roles for 653 circulating biomarkers in mediating risk of 3 ischemic stroke subtypes (large artery atherosclerosis, cardioembolic stroke, and small artery occlusion). Second, for the 7 identified biomarkers found to be significantly associated with risk of at least 1 ischemic subtype, we examined causal roles in mediating risk for hemorrhagic stroke subtypes (subarachnoid and intracerebral hemorrhages). Third, we explored a broad spectrum of side effects associated with targeting identified biomarkers for ischemic stroke treatment by expanding the previous analysis to 679 disease traits, belonging to 1 of 16 different International Classification of Disease–9 chapters. At each stage, a Bonferroni-corrected *P* value threshold was applied, accounting for both the number of biomarkers and the outcomes analyzed.



## Phenome-Wide MR Analysis of 679 Disease Traits

Lastly, we expanded the previous exploration of side effects to include nonstroke phenotypes by performing an agnostic phenome-wide MR (Phe-MR) analysis of 679 disease traits (Figure 1; Table III in the online-only Data Supplement). The primary purpose of the Phe-MR analysis was to elucidate potential on-target side effects (beneficial or adverse) associated with hypothetical interventions that reduced ischemic stroke risk by targeting identified biomarkers. SNP-biomarker effects were derived from the same GWAS meta-analysis as in the primary ischemic stroke analysis (N≤20 509). SNP-outcome effects were derived from the same UK Biobank cohort (N≤408 961) as used in the previous analysis but using a set of publicly available summary statistics provided by Zhou et al.[26] Disease outcomes were defined based on "PheCodes," a system developed to organize International Classification of Diseases and Related Health Problems codes into phenotypic outcomes suitable for systematic genetic analysis of numerous disease traits.[26,29] Sex-specific outcomes and outcomes with fewer than 500 cases were excluded because of the lack of data availability and statistical power, respectively. In contrast to previous results, Phe-MR findings were standardized to a change in biomarker level corresponding to a 10% reduction in ischemic stroke subtype risk (derived based on stage 1 associations). The results were standardized in this manner to (1) frame results in a clinical context (ie, what side effects might occur if identified biomarkers are therapeutically targeted for ischemic stroke?); (2) harmonize directionality among identified biomarkers that could either increase or decrease stroke risk; and (3) enable direct comparison of side effects (magnitude and directionality) between biomarkers. Also, given the inherent redundancy between PheCodes, we strived to improve interpretability of results by systematically selecting representative phenotypes. Parent disease categories were prioritized over individual disease entities when appropriate. For example, in the scenario that pulmonary heart disease and its constituents (primary pulmonary hypertension and chronic pulmonary heart disease) all surpassed Bonferroni significance, only pulmonary heart disease would be reported.

## Research Ethics Board Approval

The ORIGIN study received approvals from the local ethics committees at all participating sites, and all participants provided written informed consent. The UK Biobank study received approval from the National Health Service National Research Ethics Service North West.

# RESULTS

## Screening the Proteome for Causal Mediators of Ischemic Stroke Subtypes

Overall, 653 unique circulating biomarkers were tested for causal associations with 3 ischemic stroke sub-

types, representing 1959 distinct analyses (Bonferroni $P$=0.05/1959=2.55×10$^{-5}$). For a complete list of specific proteins analyzed, see Table IV in the online-only Data Supplement. At Bonferroni significance, MR analyses revealed causal relationships between 8 biomarker-stroke subtype pairs, encompassing 7 unique biomarkers (Table 1). Genetically determined levels of histo-blood group ABO system transferase, coagulation factor XI (F11), and LPA were associated with greater risk of ischemic stroke, whereas cluster of differentiation 40 (CD40), tumor necrosis factor–like weak inducer of apoptosis (TNFSF12), MMP12, and scavenger receptor class A5 (SCARA5) were associated with lower risk of ischemic stroke. Specifically, LPA increased the risk of large artery atherosclerosis (odds ratio [OR]=1.22; 95% CI, 1.14–1.30; $P$=3.19×10$^{-9}$), whereas CD40 (OR=0.73; 95% CI, 0.66–0.80; $P$=1.90×10$^{-10}$) and MMP12 (OR=0.83; 95% CI, 0.77–0.90; $P$=7.69×10$^{-7}$) decreased the risk of large artery atherosclerosis. ABO increased the risks of both large artery atherosclerosis (OR=1.08; 95% CI, 1.05–1.12; $P$=2.43×10$^{-7}$) and cardioembolic stroke (OR=1.06; 95% CI, 1.04–1.09; $P$=4.54×10$^{-7}$). F11 increased risk of cardioembolic stroke (OR=1.25; 95% CI, 1.16–1.36; $P$=1.34×10$^{-8}$), whereas TNFSF12 (OR=0.86; 95% CI, 0.81–0.91; $P$=7.69×10$^{-7}$) and SCARA5 (OR=0.78; 95% CI, 0.70–0.88; $P$=1.46×10$^{-5}$) were associated with decreased risk of cardioembolic stroke. No significant biomarker associations were found for small artery occlusion. MR sensitivity analyses were consistent with primary analyses, without evidence of directional pleiotropy or outliers driving associations (see Table V and Figures I–III in the online-only Data Supplement).

Although ABO was the only biomarker associated with multiple ischemic subtypes at Bonferroni significance, CD40, MMP12, and SCARA5 exhibited nominal associations ($P$<0.05) with other ischemic subtypes. All nominal associations with secondary subtypes exhibited the same direction-of-effect as the primary subtype (Figure 2) including CD40 with decreased risk of small artery occlusion (OR=0.91; 95% CI, 0.83–0.99; $P$=0.03); MMP12 with decreased risks of cardioembolic stroke (OR=0.94; 95% CI, 0.90–0.99; $P$=0.01) and small artery occlusion (OR=0.93; 95% CI, 0.89–0.99; $P$=0.01); and SCARA5 with decreased risk of small artery occlusion (OR=0.86; 95% CI, 0.76–0.87; $P$=0.02). Associations between identified biomarkers and aggregate phenotypes (all ischemic stroke and any stroke) are also present in Table VI in the online-only Data Supplement.

## Characterizing the Effects of Ischemic Stroke Biomarkers on Intracranial Bleeding

Next, we sought to investigate whether identified biomarkers mediated risk of hemorrhagic stroke subtypes (Figure 3). After multiple hypotheses testing correction, significant associations were found for MMP12, SCARA5, and TNFSF12. Genetically determined levels of MMP12 increased risk of intracerebral hemorrhage (OR=1.22; 95% CI, 1.09–1.36; $P$=4.84×10$^{-4}$), and TNFSF12 increased risks of both subarachnoid hemorrhage (OR=1.37; 95% CI, 1.20–1.56; $P$=3.12×10$^{-6}$) and intracerebral hemorrhage (OR=1.37; 95% CI, 1.18–1.59; $P$=2.84×10$^{-5}$). In contrast, SCARA5 decreased the risk of subarachnoid hemorrhage (OR=0.67; 95% CI, 0.52–0.85; $P$=0.001). For these associations, there was no evidence of imbalanced pleiotropy, and other MR methods yielded consistent effect estimates (Table VII in the online-only Data Supplement). Significant associations were not observed for either lobar or nonlobar intracerebral hemorrhage subtypes (Table VIII in the online-only Data Supplement).

**Table 1. Summary of Significant Biomarkers Representing Causal Mediators for Ischemic Stroke Subtypes**

| Biomarker | Stroke Subtype | Number of SNPs | Variance in Biomarker Levels Explained | Odds Ratio | 95% CI | *P* Value |
|---|---|---|---|---|---|---|
| CD40 | LAA | 23 | 0.06 | 0.73 | 0.66–0.80 | 1.90×10$^{-10}$ |
| LPA | LAA | 31 | 0.12 | 1.22 | 1.14–1.30 | 3.19×10$^{-9}$ |
| F11 | CES | 6 | 0.01 | 1.25 | 1.16–1.36 | 1.34×10$^{-8}$ |
| ABO | LAA | 39 | 0.30 | 1.08 | 1.05–1.12 | 2.43×10$^{-7}$ |
| ABO | CES | 39 | 0.30 | 1.06 | 1.04–1.09 | 4.54×10$^{-7}$ |
| TNFSF12 | CES | 17 | 0.05 | 0.86 | 0.81–0.91 | 7.69×10$^{-7}$ |
| MMP12 | LAA | 33 | 0.14 | 0.83 | 0.77–0.90 | 6.56×10$^{-6}$ |
| SCARA5 | CES | 14 | 0.02 | 0.78 | 0.70–0.88 | 1.46×10$^{-5}$ |

All displayed results surpassed correction for multiple hypotheses testing ($P$<2.55×10$^{-5}$). Odds ratios are expressed in terms of risk per 1–standard deviation increase in biomarker levels. ABO indicates histo-blood group ABO system transferase; CD40, cluster of differentiation 40; CES, cardioembolic stroke; F11, coagulation factor XI; LAA, large artery atherosclerosis; LPA, apolipoprotein(a); MMP12, matrix metalloproteinase-12; SCARA5, scavenger receptor class A5; SNP, single-nucleotide polymorphism; and TNFSF12, tumor necrosis factor–like weak inducer of apoptosis.

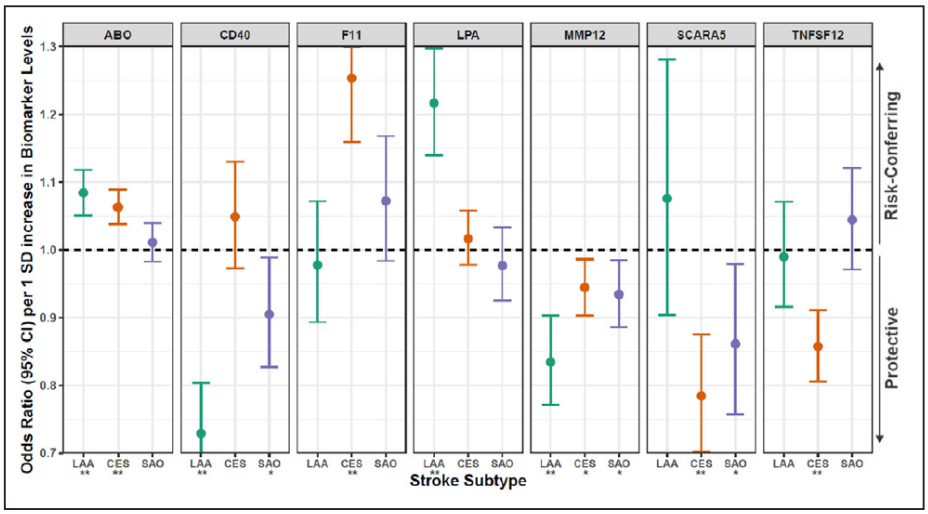



**Figure 2. Association between identified biomarkers and risk for ischemic stroke subtypes.**
Associations above the black midline represent risk-conferring effects, and those below the black midline represents protective effects. CI values are truncated at odds ratios greater than 1.30 and less than 0.70. *Nominally significant ($P$<0.05). **Bonferroni significant ($P$<0.05/(653×3)=2.55×10$^{-5}$). ABO indicates histo-blood group ABO system transferase; CD40, cluster of differentiation 40; CES, cardioembolic stroke; F11, coagulation factor XI; LAA, large artery atherosclerosis; LPA, apolipoprotein(a); MMP12, matrix metalloproteinase-12; SAO, small artery occlusion; SCARA5, scavenger receptor class A5; and TNFSF12, tumor necrosis factor–like weak inducer of apoptosis.

## Phe-MR Analysis of Identified Biomarkers for Risk of 679 Diseases

After identifying roles for some ischemic stroke biomarkers in intracranial bleeding, we conducted a broader analysis of 679 disease traits to more comprehensively depict side-effect profiles for each biomarker. Unlike the previous MR analyses, Phe-MR results were standardized to a 10% reduction in ischemic stroke subtype risk mediated through a given biomarker. As such, resultant associations can be interpreted as concomitant side effects

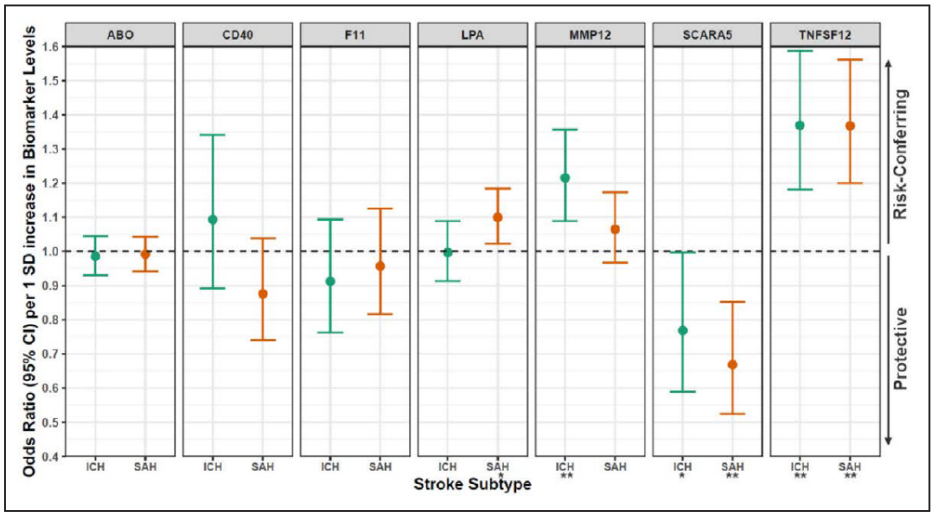



**Figure 3. Association between identified biomarkers and risk for hemorrhagic stroke subtypes.**
Associations above the black midline represent risk-conferring effects, and those below the black midline represents protective effects. *Nominally significant ($P$<0.05). **Bonferroni significant ($P$<0.05/(7×2)=0.004). ABO indicates histo-blood group ABO system transferase; CD40, cluster of differentiation 40; F11, coagulation factor XI; ICH, intracerebral hemorrhage; LPA, apolipoprotein(a); MMP12, matrix metalloproteinase-12; SAH, subarachnoid hemorrhage; SCARA5, scavenger receptor class A5; and TNFSF12, tumor necrosis factor–like weak inducer of apoptosis.

expected to arise if each biomarker is targeted therapeutically. Overall, 71 significant associations were identified (Table 2; Figure 4; Figure IV in the online-only Data Supplement), of which 53 (75%) exhibited beneficial side effects. This enrichment in beneficial vs. deleterious side effects was statistically significant ($P$=6.26×10$^{-6}$).

Examining individual biomarkers, 6 of the 7 biomarkers were associated with multiple diseases except for SCARA5, which was associated with none (Figure 4). ABO was associated with the most disease traits (28), followed by LPA (21), TNFSF12 (11), CD40 (5), F11 (3), and MMP12 (3). All disease associations with F11, LPA, and MMP12 were protective, followed by a lower proportion for CD40 (80%), ABO (71%), and TNFSF12 (18%). Grouping individual phenotypes based on disease category, the circulatory system was most commonly implicated for ABO, F11, LPA, and TNSF12 (Table 2). The most significant disease association for each biomarker was phlebitis or thrombophlebitis for ABO (OR=0.67; 95% CI, 0.62–0.71; $P$=1.36×10$^{-29}$), umbilical hernia for CD40 (OR=0.93; 95% CI, 0.90–0.96; $P$=1.07×10$^{-6}$), pulmonary heart disease for F11 (OR=0.82; 95% CI, 0.78–0.86; $P$=2.90×10$^{-16}$), ischemic heart disease for LPA (OR=0.89; 95% CI, 0.88–0.90; $P$=9.18×10$^{-110}$), cardiac shunt/heart septal defect for MMP12 (OR=0.82; 95% CI, 0.76–0.88; $P$=1.02×10$^{-7}$), and hypertension for TNFSF12 (OR=1.06; 95% CI, 1.04–1.07; $P$=1.39×10$^{-15}$). Sensitivity analyses expressed in terms of SD biomarker increase are presented in Table IX in the online-only Data Supplement.

## DISCUSSION

In this report, we used MR to systematically screen 653 circulating proteins and identified causal mediators for large artery atherosclerosis (ABO, CD40, LPA, and MMP12) and cardioembolic stroke (ABO, F11, SCARA5, and TNFSF12). MR was further applied to predict on-target side effects associated with potential ischemic stroke treatment via intervention of identified biomarkers. In addition to ischemic stroke, MMP12, SCARA5, and TNFSF12 were found to influence risk of hemorrhagic stroke subtypes. Beyond stroke, all identified biomarkers except for SCARA5 were found to mediate the risk of multiple nonstroke disorders.

The primary MR analysis revealed 7 causal mediators for ischemic stroke, and 5 of these have been previously implicated by genetic studies.[16,18,30–32] F11 is a coagulation factor whose inhibition has been hypothesized to reduce thrombosis without increasing bleeding tendency.[33] In our study, genetically determined F11 levels increased the risk of cardioembolic stroke without affecting risk of subarachnoid or intracerebral hemorrhage ($P$>0.10; Table VII in the online-only Data Supplement). This is consistent with initial RCT findings demonstrating less bleeding in 150 patients assigned to F11 inhibition vs. enoxaparin.[34] A phase II RCT for secondary ischemic stroke prevention is currently planned (NCT03766581). LPA is a well established causal risk factor for cardiovascular disease.[16] Phe-MR analysis predicted LPA inhibition to be safe with few adverse side effects and to be particularly effective for large artery atherosclerosis prevention. Indeed, a potent LPA inhibitor elicited no safety concerns in a phase II trial and will be tested for secondary prevention (NCT03070782). ABO is a glycosyltransferase enzyme whose activity determines ABO blood type. Consistent with our findings, epidemiological studies suggest that individuals with O blood type, devoid of any enzymatic activity, are protected from thrombosis.[35] However, therapies targeting ABO may be complicated given their uncertain effect on blood type and numerous side effects predicted by Phe-MR. CD40 is a member of the tumor necrosis factor receptor superfamily, and MMP12

**Table 2. Descriptive Summary of Significant Phenome-Wide Mendelian Randomization Findings Representing On-Target Side Effects of Biomarker Intervention**

| Biomarker | Number of Significant Disease Associations | Number of Beneficial/ Deleterious Side Effects (%) | Top ICD-9 Disease Chapters (%) | Most Significant Disease (Beneficial or Deleterious) |
|---|---|---|---|---|
| ABO | 28 | 20/8 (71) | Circulatory system (32); digestive (29) | Phlebitis and thrombophlebitis (B) |
| CD40 | 5 | 4/1 (80) | Respiratory (40) | Umbilical hernia (B) |
| F11 | 3 | 3/0 (100) | Circulatory system (100) | Pulmonary heart disease (B) |
| LPA | 21 | 21/0 (100) | Circulatory system (67) | Ischemic heart disease (B) |
| MMP12 | 3 | 3/0 (100) | Sense organs (33); congenital anomalies (33); injuries and poisonings (33) | Cardiac shunt/heart septal defect (B) |
| SCARA5 | 0 | — | — | — |
| TNFSF12 | 11 | 2/9 (18) | Circulatory system (55); digestive (27) | Hypertension (D) |
| Total | 71 | 53/18 (75) | Circulatory system (45) | — |

Reported in this table are the number of significant disease associations ($P$<1.07×10$^{-5}$), the number of beneficial and deleterious side effects, the ICD-9 disease chapters comprising greater than 20% of significant associations, and the most significant disease association for each biomarker. ABO indicates histo-blood group ABO system transferase; B, beneficial; CD40, cluster of differentiation 40; D, deleterious; F11, coagulation factor XI; ICD, International Classification of Disease; LPA, apolipoprotein(a); MMP12, matrix metalloproteinase-12; SCARA5, scavenger receptor class A5; and TNFSF12, tumor necrosis factor–like weak inducer of apoptosis.

ORIGINAL RESEARCH ARTICLE

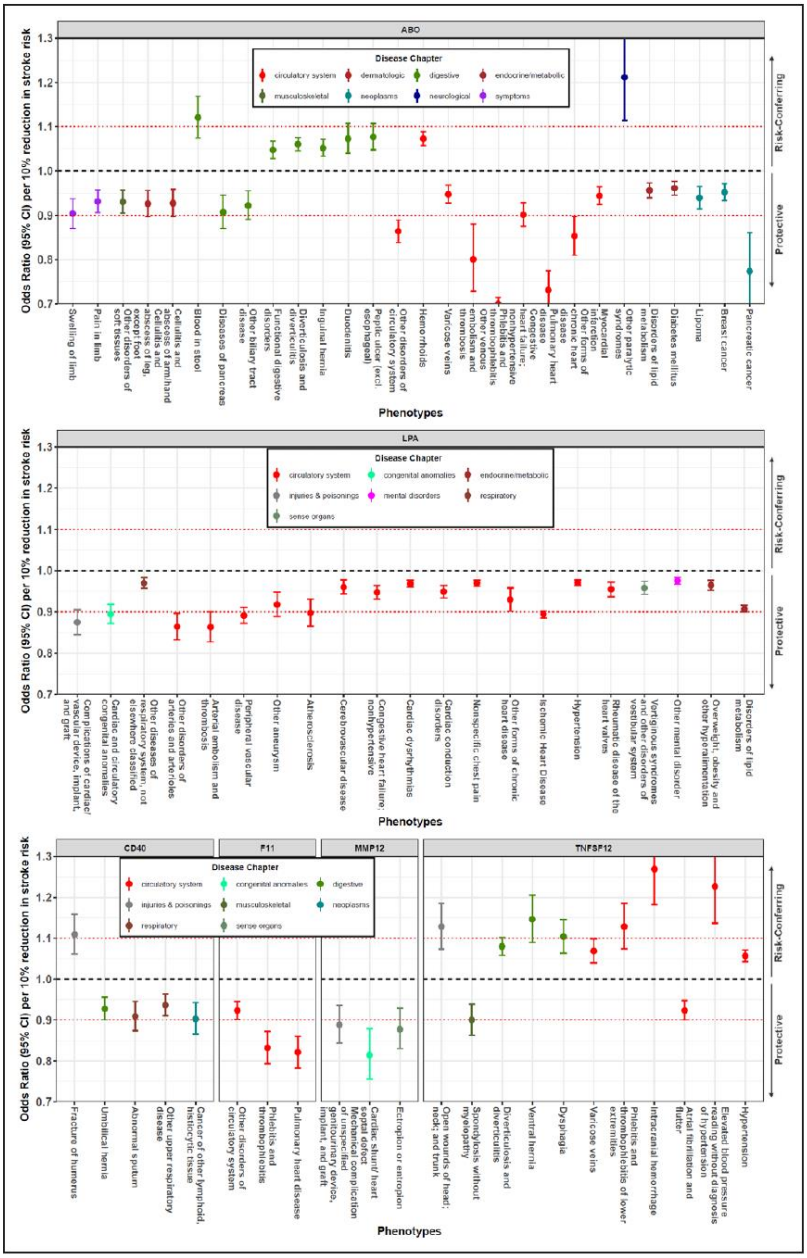



**Figure 4. Potential on-target side effects associated with biomarker intervention revealed by Phe–Mendelian randomization analysis.**
Only Bonferroni-significant disease associations are illustrated ($P<0.05/(7\times697)=1.07\times10^{-5}$). Simply, the results can be perceived as on-target side effects for a hypothetical drug that reduces risk of a given ischemic stroke subtype by 10% through intervention of circulating biomarker levels. Specifically, risk of disease is expressed per 10% reduction in risk for the specific ischemic subtype that each biomarker was associated with in the primary Mendelian randomization analysis. For ABO, which was associated with 2 ischemic stroke subtypes (large artery atherosclerosis and cardioembolic stroke), the results were standardized to a 10% reduction in risk of large artery atherosclerosis. Associations above the horizontal black midline represent deleterious side effects. Conversely, associations below this line represent beneficial side effects. The horizontal red line (odds ratio=1.10) represents the point at which decreased ischemic stroke risk is counterbalanced by an equal increase in disease risk. ABO indicates histo-blood group ABO system transferase; CD40, cluster of differentiation 40; F11, coagulation factor XI; LPA, apolipoprotein(a); MMP12, matrix metalloproteinase-12; and TNFSF12, tumor necrosis factor–like weak inducer of apoptosis.

is a member of the matrix metalloproteinase family that contributes to vascular remodeling. For both biomarkers, the relationship between circulating biomarker levels and cardiovascular disease risk remains uncertain because studies report conflicting direction-of-effects.[18,32,36] In our study, genetically determined circulating levels of CD40 and MMP12 were protective for large artery atherosclerosis. Future studies are warranted to clarify the directionality of these associations.

SCARA5 and TNFSF12 represent novel, protective mediators for cardioembolic stroke. SCARA5 is a scavenger receptor that exports ferritin-bound iron from circulation to parenchymal tissues, including the heart and brain.[37] Along with a recent MR analysis suggesting that genetically lower serum iron decreases cardioembolic stroke risk, our findings indicate a broader role for iron homeostasis in cardioembolic stroke pathogenesis.[38] SCARA5 significantly reduced risk for subarachnoid hemorrhage, and a consistent albeit nonsignificant effect was observed for intracerebral hemorrhage. Notably, acute administration of iron chelators is currently being tested in hemorrhagic stroke patients to improve recovery (NCT02175225 and NCT02875262). The utility of iron chelation for secondary prevention of hemorrhagic stroke is currently unknown, although iron restriction has been shown to attenuate aneurysm formation in mice.[39] Additionally, Phe-MR analysis did not reveal any deleterious side effects for SCARA5, and the current literature supports its safety, because SCARA5 has been reported to be protective for multiple cancers and retinopathy.[40–42] Overall, SCARA5 represents a promising therapeutic target for cardioembolic stroke and subarachnoid hemorrhage, and further research is warranted to decipher the mechanisms through which it protects against both ischemic and hemorrhagic strokes.

TNFSF12 encodes for tumor necrosis factor–like weak inducer of apoptosis, a pleiotropic cytokine that regulates numerous biological processes through binding of its receptor, tumor necrosis factor receptor superfamily member 12A (TNFRSF12A).[43] We found TNFSF12 to be protective against cardioembolic stroke, and Phe-MR results suggested that this may be mediated by lower risk for atrial fibrillation. Indeed, TNFSF12 is a well established marker of atrial fibrillation as corroborated by genetic, epidemiological, and animal investigations.[44,45] Beyond atrial fibrillation, there are also indications that TNFSF12 plays a distinct role in the brain's response to stroke. Serum TNFSF12 concentrations have been shown to be higher in stroke patients than stroke-free individuals.[46] Additionally, in mice, poststroke administration of a decoy receptor for TNFSF12 reduces infarct sizes, prevents blood–brain barrier leakage, and improves functional outcomes.[46,47] Consistent with these results, we found that genetically determined levels of TNFSF12 significantly increased risk for intracerebral and subarachnoid hemorrhages. As such, potential adverse side effects on intracranial bleeding should be considered in gauging the therapeutic utility of TNFSF12 for cardioembolic stroke treatment. Notably, TNFSF12 has been reported to interact with cluster of differentiation 163 (CD163), a scavenger receptor for circulating hemoglobin.[48] In an intracerebral hemorrhage mouse model, CD163 mitigated hematoma expansion by reducing free circulating hemoglobin.[49] This protective effect may be partially mediated by reduced TNFSF12 signaling because CD163 is known to act as a decoy receptor for circulating TNFSF12. Together, SCARA5 and TNFSF12 implicate iron and hemoglobin sequestration as potential therapeutic pathways for cardioembolic stroke and hemorrhagic stroke treatment.

Study findings have several implications. First, many additional drug targets may be uncovered as biomarker testing becomes more comprehensive. Although we evaluated causal roles for 653 proteins, the most by any stroke MR study to date, these only represent ~3% of all known proteins. Second, a subtype-specific approach was found to be effective for identifying novel drug targets because most identified biomarkers generally exhibited strong specificity for a single subtype. In fact, had all-cause ischemic stroke been used as the primary outcome, 3 biomarkers would not have been identified at Bonferroni significance. Consequently, more detailed phenotyping within ischemic subtypes (eg, lipohyalinosis vs. branch atheromatous small artery occlusion) may also enhance drug target discovery. Third, Phe-MR results suggest that biomarkers affecting stroke may also mediate the risk of many nonstroke diseases including circulatory and noncirculatory disorders. Auspiciously, most (75%) predicted side effects were beneficial and not deleterious, suggesting that if drugs were developed to treat ischemic stroke through the identified biomarkers, most side effects would be beneficial. Fourth, Phe-MR analysis serves as a powerful preclinical tool for drug target prioritization and to help anticipate target-mediated side effects. Further investigation of identified biomarkers is necessary, and study of associated proteins may aid in the identification of more direct mediators that are better targets in terms of both effectiveness and safety.

The interpretation and generalizability of study findings are limited by several factors. The following considerations relate to the derivation of genetic instruments and thus generally apply to all analyses conducted. First, specificity of protein quantification may be reduced in the presence of genetic variation and qualitative protein changes (eg, amino acid substitutions, post-translational modifications, splice variations). Qualitative changes may unpredictably affect binding affinity between detection molecules (eg, aptamers or antibodies) and protein targets leading to misinterpretation of qualitative effects as quantitative.[50] Granted, the inclusion of studies incorporating different protein quantification technologies and the use of multiple independent genetic instruments

for each biomarker reduce the likelihood that reported associations are driven by technology-specific artefacts. Second, most study participants included in the current analyses were of European ancestry because of limitations in data availability. Future biomarker GWAS in non-Europeans are necessary to enable transethnic MR analyses, which are expected to lead to more generalizable findings.[21] Third, to mitigate the possibility that genetic instruments acted through alternative pathways (ie, horizontal pleiotropy), we excluded variants distal to the encoding gene, but this approach may also exclude valid genetic instruments as in the work of Georgakis et al.[17] Conversely, stroke-associated biomarkers lacking causal support from previous MR studies (ie, CHI3L1, CRP, and CST3) were also nonsignificant ($P>0.05$) in our analyses, which supports the specificity of this approach (Table X in the online-only Data Supplement). A limitation of the primary analysis of ischemic stroke subtypes was the inability to replicate results in the UK Biobank study because of incompatibility of subtype classification systems (ie, TOAST vs. International Classification of Diseases and Related Health Problems) and sparse subtyping. For similar reasons, subtypes of intracerebral hemorrhage were not investigated in the second stage analysis within UK Biobank, despite evidence for distinct genetic architectures between lobar and nonlobar hemorrhages by Woo et al.[28] Accordingly, future studies should address whether the observed TNFSF12 and MMP12 associations with intracerebral hemorrhage are subtype-specific. Lastly, although the PheCode system enables systematic assessment of a many disease traits, there are several shortcomings. Because PheCodes rely on hospital diagnoses, diseases less likely to result in hospital admission may be poorly represented. Moreover, validity of PheCode-defined diagnoses as compared with the gold standard is unknown for many traits.

## Conclusions

Systematic MR analysis of the circulating proteome revealed 7 causal mediators for ischemic stroke. SCARA5 and TNFSF12 represent 2 novel protective biomarkers, which also affect risk for intracranial bleeding. Phe-MR analysis suggests important roles for stroke biomarkers in many other disorders, involving both cardiovascular and noncardiovascular systems. LPA, F11, and SCARA5 are particularly attractive drug targets with no adverse side effects predicted. Further research is warranted to assess the viability of these protein biomarkers as drug targets for stroke treatment.

## ARTICLE INFORMATION

Received February 10, 2019; accepted June 3, 2019.

The online-only Data Supplement is available with this article at https://www.ahajournals.org/doi/suppl/10.1161/circulationaha.119.040180.

### Correspondence

Guillaume Paré, MD, MSc, 237 Barton St East, Rm C4-126, Hamilton, Ontario L8L 2X2, Canada. Email pareg@mcmaster.ca

### Affiliations

Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Thrombosis and Atherosclerosis Research Institute, Hamilton Health Sciences, Ontario, Canada (M.C., J.S., M.P., P.M.-S., R.L., A.S., H.C.G., G.P.). Departments of Biochemistry (M.C., G.P.), Medical Sciences (P.M.-S.), Clinical Epidemiology and Biostatistics (H.C.G., G.P.), Pathology and Molecular Medicine (G.P.), and Medicine, Division of Neurology, McMaster University, Hamilton, Ontario, Canada (A.S.).

### Acknowledgments

The authors acknowledge the important contributions of the many publicly available data sets used in this report's analyses including the UK Biobank (application 15255, "Identification of the shared biological and sociodemographic factors underlying cardiovascular disease and dementia risk"), MEGASTROKE (International Stroke Genetics Consortium), Stroke Genetics Network (SiGN), YFS/FINRISK, IMPROVE, KORA F4/QMDiab, and INTERVAL.

### Sources of Funding

None.

### Disclosures

Dr Paré has received consulting fees from Sanofi, Bristol–Myers Squibb, Lexicomp, and Amgen; has received support for research through his institution from Sanofi; and has received support from the Canada Research Chair in Genetic and Molecular Epidemiology, CISCO Professorship in Integrated Health Systems. Dr Gerstein holds the McMaster–Sanofi Population Health Institute Chair in Diabetes Research and Care; has received research grant support from AstraZeneca, Eli Lilly, Merck, Novo Nordisk, and Sanofi; has received honoraria for speaking from AstraZeneca, Boehringer Ingelheim, Eli Lilly, Novo Nordisk, and Sanofi; and has received consulting fees from Abbott, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Merck, Novo Nordisk, Janssen, Cirius, Kowa, and Sanofi. Dr Shoamanesh is part of the steering committee for ANNEXA-1; has received honoraria from Bayer AG and ServierCanada Inc; has received grant support from Bayer AG, Bristol–Myers Squibb, and ServierCanada Inc; and has received consulting fees from Daiichi Sankyo, Bayer AG, Boehringer Ingelheim, Bristol–Myers Squibb, ServierCanada Inc, and ApoPharma Inc. The other authors report no conflicts.

## REFERENCES

1. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, Ahmed M, Aksut B, Alam T, Alam K, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *J Am Coll Cardiol*. 2017;70:1–25. doi: 10.1016/j.jacc.2017.04.052
2. SPS3 Study Group, Benavente OR, Coffey CS, Conwit R, Hart RG, McClure LA, Pearce LA, Pergola PE, Szychowski JM. Blood-pressure targets in patients with recent lacunar stroke: the SPS3 randomised trial. *Lancet*. 2013;382:507–515. doi: 10.1016/S0140-6736(13)60852-1
3. Dwyer DJ, Camacho DM, Kohanski MA, Callura JM, Collins JJ. Stenting versus aggressive medical therapy for intracranial arterial stenosis. *N Engl J Med*. 2013;46:561–572. doi: 10.1056/NEJMoa1105335
4. Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE 3rd. Classification of subtype of acute ischemic stroke: definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24:35–41. doi: 10.1161/01.str.24.1.35
5. Prins BP, Abbasi A, Wong A, Vaez A, Nolte I, Franceschini N, Stuart PE, Guterriez Achury J, Mistry V, Bradfield JP, et al; PAGE Consortium; International Stroke Genetics Consortium; Systemic Sclerosis consortium; Treat OA consortium; DIAGRAM Consortium; CARDIoGRAMplusC4D Consortium; ALS consortium; International Parkinson's Disease Genomics Consortium; Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium; CKDGen consortium; GERAD1 Consortium; International Consortium for Blood Pressure; Schizophrenia Working Group of the Psychiatric Genomics Consortium; Inflammation Working Group of the CHARGE Consortium. Investigating the causal relationship of C-

reactive protein with 32 complex somatic and psychiatric outcomes: a large-scale cross-consortium Mendelian randomization study. *PLoS Med*. 2016;13:e1001976. doi: 10.1371/journal.pmed.1001976
6. van der Laan SW, Fall T, Soumaré A, Teumer A, Sedaghat S, Baumert J, Zabaneh D, van Setten J, Isgum I, Galesloot TE, et al. Cystatin C and cardiovascular disease: a Mendelian randomization study. *J Am Coll Cardiol*. 2016;165:255–269. doi: 10.1016/j.jacc.2016.05.092r
7. Kjaergaard AD, Johansen JS, Bojesen SE, Nordestgaard BG. Elevated plasma YKL-40, lipids and lipoproteins, and ischemic vascular disease in the general population. *Stroke*. 2015;46:329–335. doi: 10.1161/STROKEAHA.114.007657
8. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47:856–860. doi: 10.1038/ng.3314
9. Ahola-Olli AV, Würtz P, Havulinna AS, Aalto K, Pitkänen N, Lehtimäki T, Kähönen M, Lyytikäinen LP, Raitoharju E, Seppälä I, et al. Genome-wide association study identifies 27 loci influencing concentrations of circulating cytokines and growth factors. *Am J Hum Genet*. 2017;100:40–50. doi: 10.1016/j.ajhg.2016.11.007
10. Folkersen L, Fauman E, Sabater-Lleal M, Strawbridge RJ, Frånberg M, Sennblad B, Baldassarre D, Veglia F, Humphries SE, Rauramaa R, et al; IMPROVE study group. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet*. 2017;13:e1006706. doi: 10.1371/journal.pgen.1006706
11. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, Sarwath H, Thareja G, Wahl A, DeLisle RK, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun*. 2017;8:14357. doi: 10.1038/ncomms14357
12. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018;558:73–79. doi: 10.1038/s41586-018-0175-2
13. Sjaarda J, Gerstein H, Chong M, Yusuf S, Meyre D, Anand SS, Hess S, Paré G. Blood CSF1 and CXCL12 as causal mediators of coronary artery disease. *J Am Coll Cardiol*. 2018;72:300–310. doi: 10.1016/j.jacc.2018.04.067
14. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23:R89–R98. doi: 10.1093/hmg/ddu328
15. Bennett DA, Holmes MV. Mendelian randomisation in cardiovascular research: an introduction for clinicians. *Heart*. 2017;103:1400–1407. doi: 10.1136/heartjnl-2016-310605
16. Helgadottir A, Gretarsdottir S, Thorleifsson G, Holm H, Patel RS, Gudnason T, Jones GT, van Rij AM, Eapen DJ, Baas AF, et al. Apolipoprotein(a) genetic sequence variants associated with systemic atherosclerosis and coronary atherosclerotic burden but not with venous thromboembolism. *J Am Coll Cardiol*. 2012;60:722–729. doi: 10.1016/j.jacc.2012.01.078
17. Georgakis MK, Gill D, Rannikmäe K, Traylor M, Anderson CD, Lee JM, Kamatani Y, Hopewell JC, Worrall BB, Bernhagen J, et al. Genetically determined levels of circulating cytokines and risk of stroke. *Circulation*. 2019;139:256–268. doi: 10.1161/CIRCULATIONAHA.118.035905
18. Traylor M, Mäkelä KM, Kilarski LL, Holliday EG, Devan WJ, Nalls MA, Wiggins KL, Zhao W, Cheng YC, Achterberg S, et al; METASTROKE, International Stroke Genetics Consortium, Wellcome Trust Case Consortium 2 (WTCCC2). A novel MMP12 locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach. *PLoS Genet*. 2014;10:e1004469. doi: 10.1371/journal.pgen.1004469
19. Rao AS, Lindholm D, Rivas MA, Knowles JW, Montgomery SB, Ingelsson E. Large-scale phenome-wide association study of PCSK9 variants demonstrates protection against ischemic stroke. *Circ Genom Precis Med*. 2018;11:e002162. doi: 10.1161/CIRCGEN.118.002162
20. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–2191. doi: 10.1093/bioinformatics/btq340
21. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese AK, van der Laan SW, Gretarsdottir S, et al; AFGen Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium; International Genomics of Blood Pressure (iGEN-BP) Consortium; INVENT Consortium; STARNET; BioBank Japan Cooperative Hospital Group; COMPASS Consortium; EPIC–CVD Consortium; EPIC-InterAct Consortium; International Stroke Genetics Consortium (ISGC); METASTROKE Consortium; Neurology Working Group of the CHARGE Consortium; NINDS Stroke Genetics Network (SiGN); UK Young Lacunar DNA Study; MEGASTROKE Consortium. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524–537. doi: 10.1038/s41588-018-0058-3
22. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74. doi: 10.1038/nature15393
23. Zhao Q, Wang J, Hemani G, Bowden J, Small D. Statistical inference in two-sample summary data Mendelian randomization using robust adjusted profile score. *arXiv*. 2019. Available from: https://arxiv.org/abs/1801.09652. arXiv ID: 1801.09652v3. Accessed January 28, 2019.
24. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408. doi: 10.7554/eLife.34408
25. Sun L, Clarke R, Bennett D, Guo Y, Walters RG, Hill M, Parish S, Millwood IY, Bian Z, Chen Y, et al; China Kadoorie Biobank Collaborative Group; International Steering Committee; International Co-ordinating Centre, Oxford; National Co-ordinating Centre, Beijing; Regional Co-ordinating Centres. Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults. *Nat Med*. 2019;25:569–574. doi: 10.1038/s41591-019-0366-x
26. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, LeFaive J, VandeHaar P, Gagliano SA, Gifford A, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50:1335–1341. doi: 10.1038/s41588-018-0184-y
27. Schnier C, Denaxas S, Eggo R, Patel R, Zhang Q, Woodfield R, Flaig R, Hemingway H, Sudlow C. Identification and validation of myocardial infarction and stroke outcomes at scale in UK Biobank. *Int J Popul Data Sci*. 2017;10:1–6. doi: 10.3923/ijp.2015.327.334
28. Woo D, Falcone GJ, Devan WJ, Brown WM, Biffi A, Howard TD, Anderson CD, Brouwers HB, Valant V, Battey TW, et al; International Stroke Genetics Consortium. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am J Hum Genet*. 2014;94:511–521. doi: 10.1016/j.ajhg.2014.02.012
29. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;31:1102–1110. doi: 10.1038/nbt.2749
30. Gill D, Georgakis MK, Laffan M, Sabater-Lleal M, Malik R, Tzoulaki I, Veltkamp R, Dehghan A. Genetically determined FXI (Factor XI) levels and risk of stroke. *Stroke*. 2018;49:2761–2763. doi: 10.1161/STROKEAHA.118.022792
31. Williams FM, Carter AM, Hysi PG, Surdulescu G, Hodgkiss D, Soranzo N, Traylor M, Bevan S, Dichgans M, Rothwell PM, et al; EuroCLOT Investigators; Wellcome Trust Case Control Consortium 2; MOnica Risk, Genetics, Archiving and Monograph; MetaStroke; International Stroke Genetics Consortium. Ischemic stroke is associated with the ABO locus: the EuroCLOT study. *Ann Neurol*. 2013;73:16–31. doi: 10.1002/ana.23838
32. Huang HT, Guo J, Xiang Y, Chen JM, Luo HC, Meng LQ, Wei YS. A SNP in 5′ untranslated region of CD40 gene is associated with an increased risk of ischemic stroke in a Chinese population: a case-control study. *Genet Mol Biol*. 2017;40:442–449. doi: 10.1590/1678-4685-GMB-2016-0212
33. Zhang H, Löwenberg EC, Crosby JR, MacLeod AR, Zhao C, Gao D, Black C, Revenko AS, Meijers JC, Stroes ES, et al. Inhibition of the intrinsic coagulation pathway factor XI by antisense oligonucleotides: a novel antithrombotic strategy with lowered bleeding risk. *Blood*. 2010;116:4684–4692. doi: 10.1182/blood-2010-04-277798
34. Büller HR, Bethune C, Bhanot S, Gailani D, Monia BP, Raskob GE, Segers A, Verhamme P, Weitz JI; FXI-ASO TKA Investigators. Factor XI antisense oligonucleotide for prevention of venous thrombosis. *N Engl J Med*. 2015;372:232–240. doi: 10.1056/NEJMoa1405760
35. Wu O, Bayoumi N, Vickers MA, Clark P. ABO(H) blood groups and vascular disease: a systematic review and meta-analysis. *J Thromb Haemost*. 2008;6:62–69. doi: 10.1111/j.1538-7836.2007.02818.x
36. Didangelos A, Yin X, Mandal K, Saje A, Smith A, Xu Q, Jahangiri M, Mayr M. Extracellular matrix composition and remodeling in human abdominal aortic aneurysms: a proteomics approach. *Mol Cell Proteomics*. 2011;10:M111.008128. doi: 10.1074/mcp.M111.008128
37. Li JY, Paragas N, Ned RM, Qiu A, Viltard M, Leete T, Drexler IR, Chen X, Sanna-Cherchi S, Mohammed F, et al. Scara5 is a ferritin receptor mediating non-transferrin iron delivery. *Dev Cell*. 2009;16:35–46. doi: 10.1016/j.devcel.2008.12.002

38. Gill D, Monori G, Tzoulaki I, Dehghan A. Iron status and risk of stroke. *Stroke.* 2018;49:2815–2821. doi: 10.1161/STROKEAHA.118.022701
39. Sawada H, Hao H, Naito Y, Oboshi M, Hirotani S, Mitsuno M, Miyamoto Y, Hirota S, Masuyama T. Aortic iron overload with oxidative stress and inflammation in human and murine abdominal aortic aneurysm. *Arterioscler Thromb Vasc Biol.* 2015;35:1507–1514. doi: 10.1161/ATVBAHA.115.305586
40. Ulker D, Ersoy YE, Gucin Z, Muslumanoglu M, Buyru N. Downregulation of SCARA5 may contribute to breast cancer via promoter hypermethylation. *Gene.* 2018;673:102–106. doi: 10.1016/j.gene.2018.06.036
41. Wen X, Wang N, Zhang F, Dong C. Overexpression of SCARA5 inhibits tumor proliferation and invasion in osteosarcoma via suppression of the FAK signaling pathway. *Mol Med Rep.* 2016;13:2885–2891. doi: 10.3892/mmr.2016.4857
42. Mendes-Jorge L, Ramos D, Valença A, López-Luppo M, Pires VM, Catita J, Nacher V, Navarro M, Carretero A, Rodriguez-Baeza A, et al. L-Ferritin binding to scara5: a new iron traffic pathway potentially implicated in retinopathy. *PLoS One.* 2014;9:e106974. doi: 10.1371/journal.pone.0106974
43. Ikner A, Ashkenazi A. TWEAK induces apoptosis through a death-signaling complex comprising receptor-interacting protein 1 (RIP1), Fas-associated death domain (FADD), and caspase-8. *J Biol Chem.* 2011;286:21546–21554. doi: 10.1074/jbc.M110.203745
44. Hao L, Ren M, Rong B, Xie F, Lin MJ, Zhao YC, Yue X, Han WQ, Zhong JQ. TWEAK/Fn14 mediates atrial-derived HL-1 myocytes hypertrophy via JAK2/STAT3 signalling pathway. *J Cell Mol Med.* 2018;22:4344–4353. doi: 10.1111/jcmm.13724
45. Roselli C, Chaffin MD, Weng LC, Aeschbacher S, Ahlberg G, Albert CM, Almgren P, Alonso A, Anderson CD, Aragam KG, et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet.* 2018;50:1225–1233. doi: 10.1038/s41588-018-0133-9
46. Inta I, Frauenknecht K, Dörr H, Kohlhof P, Rabsilber T, Auffarth GU, Burkly L, Mittelbronn M, Hahm K, Sommer C, et al. Induction of the cytokine TWEAK and its receptor Fn14 in ischemic stroke. *J Neurol Sci.* 2008;275:117–120. doi: 10.1016/j.jns.2008.08.005
47. Zhang X, Winkles JA, Gongora MC, Polavarapu R, Michaelson JS, Hahm K, Burkly L, Friedman M, Li XJ, Yepes M. TWEAK-Fn14 pathway inhibition protects the integrity of the neurovascular unit during cerebral ischemia. *J Cereb Blood Flow Metab.* 2007;27:534–544. doi: 10.1038/sj.jcbfm.9600368
48. Akahori H, Karmali V, Polavarapu R, Lyle AN, Weiss D, Shin E, Husain A, Naqvi N, Van Dam R, Habib A, et al. CD163 interacts with TWEAK to regulate tissue regeneration after ischaemic injury. *Nat Commun.* 2015;6:7792. doi: 10.1038/ncomms8792
49. Roy-O'Reilly M, Zhu L, Atadja L, Torres G, Aronowski J, McCullough L, Edwards NJ. Soluble CD163 in intracerebral hemorrhage: biomarker for perihematomal edema. *Ann Clin Transl Neurol.* 2017;4:793–800. doi: 10.1002/acn3.485
50. Joshi A, Mayr M. In aptamers they trust: the caveats of the SOMAscan biomarker discovery platform from SomaLogic. *Circulation.* 2018;138:2482–2485. doi: 10.1161/CIRCULATIONAHA.118.036823

# CHAPTER 4:

## GWAS and ExWAS of blood Mitochondrial DNA copy number identifies 71 loci and highlights a potential causal role in dementia

Submitted to *eLife.* (June 1st, 2021)

**GWAS and ExWAS of blood Mitochondrial DNA copy number identifies 71 loci and highlights a potential causal role in dementia**

**Authors:** Michael Chong[1,2,3,4], Pedrum Mohammadi-Shemirani[1,2,4], Nicolas Perrot[1,2], Walter Nelson[5], Robert W. Morton[1,2,4], Sukrit Narula[1,2,6], Ricky Lali[1,2,6], Irfan Khan[1,2,4], Mohammad Khan[1,2,7], Conor Judge[1,2,8], Tafadzwa Machipisa[1,2,4,9,10], Nathan Cawte[1,2], Martin O'Donnell[1,8], Marie Pigeyre[1,2,7], Loubna Akhabir[1,2,7], Guillaume Paré[1,2,3,4,6,7]*

*Guillaume Paré. Email: pareg@mcmaster.ca

**Affiliations:**

[1]Population Health Research Institute (PHRI), David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton Health Sciences; 237 Barton Street East, Hamilton, L8L 2X2, Ontario, Canada.

[2]Thrombosis and Atherosclerosis Research Institute; 237 Barton Street East, Hamilton, L8L 2X2, Ontario, Canada.

[3]Department of Biochemistry and Biomedical Sciences, McMaster University; 1280 Main Street West, Hamilton, Ontario, L8S 4K1, Canada

[4]Department of Pathology and Molecular Medicine, Michael G. DeGroote School of Medicine, McMaster University; 1280 Main Street West, Hamilton, L8S 4K1, Ontario, Canada

[5]Centre for Data Science and Digital Health, Hamilton Health Sciences; 293 Wellington St. North, Suite 132, Hamilton, L8L 8E7, Ontario, Canada

[6]Department of Health Research Methods, Evidence, and Impact, McMaster University; 1280 Main Street West, Hamilton, L8S 4K1, Ontario, Canada

[7]Department of Medicine, McMaster University, Michael G. DeGroote School of Medicine; 1280 Main Street West, Hamilton, L8S 4K1, Ontario, Canada

[8]National University of Ireland Galway; University Road, H91 TK33, Galway, Ireland.

[9]Department of Medicine, University of Cape Town & Groote Schuur Hospital; Main Road, Observatory, Cape Town, 7925, South Africa

[10]Hatter Institute for Cardiovascular Diseases Research in Africa (HICRA) & Cape Heart Institute (CHI), Department of Medicine, University of Cape Town; 41 Chris Barnard Building, Anzio Road, Observatory, Cape Town, 7925, South Africa

**Abstract**

**Background**: Mitochondrial DNA copy number (mtDNA-CN) is an accessible blood-based measurement believed to capture underlying mitochondrial function. The specific biological processes underpinning its regulation, and whether those processes are causative for disease, is an area of active investigation.

**Methods**: We developed a novel method for array-based mtDNA-CN estimation suitable for biobank-scale studies, called "AutoMitoC". We applied AutoMitoC to 395,781 UKBiobank study participants and performed genome and exome-wide association studies, identifying novel common and rare genetic determinants. Finally, we performed two-sample Mendelian Randomization to assess whether genetically low mtDNA-CN influenced select mitochondrial phenotypes.

**Results**: Overall, genetic analyses identified 71 loci for mtDNA-CN, which implicated several genes involved in rare mtDNA depletion disorders, dNTP metabolism, and the mitochondrial central dogma. Rare variant analysis identified SAMHD1 mutation carriers

as having higher mtDNA-CN (beta=0.23 SDs; 95% CI, 0.18- 0.29; P=2.6x10-19), a potential therapeutic target for patients with mtDNA depletion disorders, but at increased risk of breast cancer (OR=1.91; 95% CI, 1.52-2.40; P=2.7x10- 8). Finally, Mendelian randomization analyses suggest a causal effect of low mtDNA-CN on dementia risk (OR=1.94 per 1 SD decrease in mtDNA-CN; 95% CI, 1.55-2.32; P=7.5x10-4). **Conclusions**: Altogether, our genetic findings indicate that mtDNA-CN is a complex biomarker reflecting specific mitochondrial processes related to mtDNA regulation, and that these processes are causally related to human diseases.

**Funding**: No funds supported this specific investigation. Awards and positions supporting authors include: Canadian Institutes of Health Research (CIHR) Frederick Banting and Charles Best Canada Graduate Scholarships Doctoral Award (MC, PM); CIHR Post-Doctoral Fellowship Award (RM); Wellcome Trust Grant number: 099313/B/12/A; Crasnow Travel Scholarship; Bongani Mayosi UCT-PHRI Scholarship 2019/2020 (TM); Wellcome Trust Health Research Board Irish Clinical Academic Training (ICAT) Programme Grant Number: 203930/B/16/Z (CJ); European Research Council COSIP Grant Number: 640580 (MO); E.J. Moran Campbell Internal Career Research Award (MP); CISCO Professorship in Integrated Health Systems and Canada Research Chair in Genetic and Molecular Epidemiology (GP)

## Introduction

Mitochondria are semi-autonomous organelles present in nearly every human cell that execute fundamental cellular processes including oxidative phosphorylation, calcium storage, and apoptotic signalling. Mitochondrial dysfunction has been implicated as the

underlying cause for many human disorders based on mechanistic *in vitro* and *in vivo* studies (Burbulla *et al.*, 2017; Desdín-micó *et al.*, 2020; Sliter *et al.*, 2020). Complementary evidence comes from recent epidemiological studies that measure mitochondrial DNA Copy Number (mtDNA-CN), a marker of mitochondrial activity that can be conveniently measured from peripheral blood. Since mitochondria contain their own unique set of genomes that are distinct from the nuclear genome, the ratio of mtDNA to nuclear DNA molecules (mtDNA-CN) in a sample serves as an accessible marker of mitochondrial quantity (Longchamps *et al.*, 2020). Indeed, observational studies suggest that individuals with lower mtDNA-CN are at higher risk of age-related complex diseases, such as coronary artery disease, sudden cardiac death, cardiomegaly, stroke, portal hypertension, and chronic kidney disease (Tin *et al.*, 2016; Ashar *et al.*, 2017; Zhang *et al.*, 2017; Hägg *et al.*, 2020). Conversely, higher mtDNA-CN levels have been associated with increased cancer incidence (Kim *et al.*, 2015; Hu, Yao and Shen, 2016).

While previous studies demonstrate that mtDNA-CN is a biomarker of mitochondrial activity associated with various diseases, evidence suggests that it may also play a direct and causative role in human health and disease. For example, in cases of mtDNA depletion syndrome, wherein rare defects in nuclear genes responsible for replicating and/or maintaining mtDNA lead to deficient mtDNA-CN (Gorman *et al.*, 2016), patients manifest with severe dysfunction of energy-dependent tissues (heart, brain, liver, and cardiac and skeletal muscles). So far, 19 genes have been reported to cause mtDNA depletion (Oyston, 1998). In addition to these rare monogenic syndromes, the importance of common genetic variation in regulating mtDNA-CN is an active area of research with

approximately 50 common loci identified so far (Cai *et al.*, 2015; Guyatt *et al.*, 2019; Longchamps, 2019; Hägg *et al.*, 2020).

To interrogate mtDNA-CN as a potential determinant of human diseases, we performed extensive genetic investigations in up to 395,781 participants from the UKBiobank study (Sudlow *et al.*, 2015). We first developed and validated a novel method for biobank-scale mtDNA-CN investigations that leverages SNP array intensities, called "AutoMitoC". Leveraging AutoMitoC-based mtDNA-CN estimates, we performed large-scale GWAS and ExWAS to identify common and rare genetic variants contributing to population-level variation in mtDNA-CN. Various analyses were then conducted to build on previous publications regarding the specific genes and pathways underlying mtDNA-CN regulation (Cai *et al.*, 2015; Guyatt *et al.*, 2019; Longchamps, 2019; Hägg *et al.*, 2020). Finally, we applied Mendelian randomization analyses to assess potential causal relationships between mtDNA-CN and disease susceptibility.

## Materials and Methods

### The UKBiobank study

The UKBiobank is a prospective cohort study including approximately 500,000 UK residents (ages 40-69 years) recruited from 2006-2010 in whom extensive genetic and phenotypic investigations have been and continue to be done (Sudlow *et al.*, 2015). All UKBiobank data reported in this manuscript were accessed through the UKBiobank data showcase under application #15525. All analyses involve the use of genetic and/or phenotypic data from consenting UKBiobank participants.

### Genetic Analysis of Common Variants

Data acquisition and quality control

Imputed genotypes (version 3) for 488,264 UKBiobank participants were downloaded through the European Genome Archive (Category 100319). Detailed sample and variant quality control are described in the supplementary methods. In special consideration of mtDNA-CN as the GWAS phenotype, we also removed variants within "NUMTs", which refer to regions of the nuclear genome that exhibit homology to the mitochondrial genome due to past transposition of mitochondrial sequences (Simone *et al.*, 2011). After quality control, 359,689 British, 10,598 Irish, 13,189 Other White, 6,172 South Asian, and 6,133 African samples had suitable array-based mtDNA-CN estimates for subsequent GWAS testing.

Association testing

GWAS were initially conducted in an ethnicity-stratified manner for common variants (MAF > 0.005). To allow for genetic relatedness between participants, GWAS were conducted using the REGENIE framework (Mbatchou *et al.*, 2020). GWAS covariates included age, age$^2$, sex, chip type, 20 genetic principal components, and blood cell counts (white blood cell, platelet, and neutrophil counts). After ethnicity-specific GWAS were performed, results were combined through meta-analysis using METAL (Willer, Li and Abecasis, 2010). European (N=383,476) and trans-ethnic (N=395,718) GWAS meta-analyses were performed. Sensitivity analyses testing for cryptic NUMT interference was conducted as per Nandakumar *et al.* (2021) (Nandakumar *et al.*, 2021). See supplementary methods for further details.

*Fine-mapping of GWAS signals*

We followed a similar protocol to Vuckovic *et al.* (2020) for fine-mapping mtDNA-CN loci (Vuckovic *et al.*, 2020). Genome-wide significant variants were consolidated into genomic blocks by grouping variants within 250kb of each other. LDstore was used to compute a pairwise LD correlation matrix for all variants within each block and across all samples included in the European GWAS meta-analysis (Benner *et al.*, 2017). For each genomic block, FINEMAP was used to perform stepwise conditional regression (Benner *et al.*, 2016). The number of conditionally independent genetic signals per genomic block was used to inform the subsequent fine-mapping search parameters. Finally, the FINEMAP random stochastic search algorithm was applied to derive 95% credible sets constituting candidate causal variants that jointly contributed to 95% (or higher) of the posterior inclusion probabilities (Benner *et al.*, 2016).

*Mitochondrial expression quantitative trait loci (mt-eQTL) and heteroplasmy look-ups*

Among GWAS hits, we searched for mt-eQTLs using information from Ali *et al*. (2019), "Nuclear genetic regulation of the human mitochondrial transcriptome"(Ali *et al.*, 2019). All variants in both Tables 1 and 2 were queried in the mtDNA-CN summary statistics. When mt-eQTLs also had reported effect estimates, the consistency in direction-of-effects between mt-eQTL and mtDNA-CN associations was reported (S2. Table 3). GWS variants associated with mean heteroplasmy levels were extracted from Nandakumar *et al.* (Nandakumar *et al.*, 2021).

*Gene prioritization & pathway analyses*

The Data-driven Expression-prioritized Integration for Complex Traits (DEPICT) v.1.1 tool was used to map mtDNA-CN loci to genes based on shared co-regulation of gene

expression using default settings (Pers *et al.*, 2015). DEPICT-prioritized genes were uploaded to the GeneMANIA web platform (https://genemania.org/). Based on the combined list of DEPICT and GeneMANIA identified genes, a network was formed in GeneMANIA using default settings. Functional enrichment analysis was then performed to identify overrepresented Gene Ontology (GO) terms among all network genes (Gene and Consortium, 2000).

*Mitochondrial annotation-based analyses*

To complement the previous pathway analyses, we labelled prioritized genes with MitoCarta3 annotations and performed subsequent statistical enrichment analyses (Rath *et al.*, 2021). MitoCarta3 is an exquisite database of mitochondrial protein annotations, which draws from mass spectrophotometry and GFP colocalization experiments of isolated mitochondria from 14 different tissues to assign all human genes statuses indicating whether the corresponding proteins are expressed in the mitochondria or not. We tested whether prioritized genes were enriched for the mitochondrial proteome by using a binomial test in R. Furthermore, a t-test was used to compare mean PGC-1A induced fold change for the subset of GWAS-prioritized genes expressed in the mitochondrial proteome as compared to the mean PGC-1A induced fold change for all 1120 nuclear MitoCarta3-annotated genes. Also, genes were categorized based on MitoCarta3 "MitoPathway" annotations.

**Genetic Analysis of Rare Variants**

*Data acquisition and quality control*

Population-level whole-exome sequencing (WES) variant genotypes (UKB data field: 23155) for 200,643 UKBiobank participants corresponding to 17,975,236 variants were downloaded using the gfetch utility. Detailed quality control of WES data is described in the supplementary methods. After quality control, 173,688 unrelated Caucasian samples remained.

*Exome-wide association study (ExWAS) to identify rare mtDNA-CN loci*

Of the 173,688 individuals, suitable AutoMitoC mtDNA-CN estimates were available for 147,740 samples. Rare variant inclusion criteria consisted of variants which were infrequent (MAF $\leq$ 0.001), non-synonymous, and predicted to be clinically deleterious by Mendelian Clinically Applicable Pathogenicity (M-CAP) v.1.4 scores (or were highly disruptive variant types including frameshift indel, stopgain, stoploss, or splicing) (Jagadeesh *et al.*, 2016). Herein, such variants are referred to as "rare variants" for simplicity. For each gene, rare allele counts were added per sample. A minor allele count of 10 was applied leading to a total of 18,890 genes analyzed (exome-wide significance $P < 0.05/18{,}890 = 2.65 \times 10^{-6}$). Linear regression was conducted using mtDNA-CN as the dependent variable and the rare alleles counts per gene as the independent variable. The same set of covariates used in the primary GWAS analysis was used in the ExWAS analysis.

*Phenome-wide association testing for rare SAMHD1 mutation carrier status*

To identify disease phenotypes associated with carrying a rare *SAMHD1* mutation, we maximized sample size for phenome-wide association testing by analyzing the larger set of 173,688 WES samples (with or without suitable mtDNA-CN estimates). Disease outcomes

were defined using the previously published "PheCode" classification scheme to aggregate ICD-10 codes from hospital episodes (field ID 41270), death registry (field ID 40001 and 40002), and cancer registry (field ID 40006) records (Denny *et al.*, 2013; Wu *et al.*, 2019). Logistic regression was applied to test the association of *SAMHD1* mutation carrier status versus 771 PheCodes (phenome-wide significance $P < 0.05/771 = 6.49 \times 10^{-5}$) with a minimal case sample size of 300 (Wei *et al.*, 2017). The same set of covariates used in the primary GWAS were also employed in this analysis.

**Mendelian Randomization Analysis**

*Disease Outcomes*

We cross-referenced a list of 36 clinical manifestations of mitochondrial disease to FinnGen consortium GWAS (release 4; November 30 2020) traits (Gorman *et al.*, 2016; Feng *et al.*, 2020). Among 2,444 FinnGen traits, 10 overlapped with mitochondrial disease and had a case prevalence greater than 1% and were chosen for two-sample Mendelian Randomization analyses. These 10 traits included type 2 diabetes (N=23,364), mood disorder (N=20,288), sensorineural hearing loss (N=12,550), cerebrovascular disease (N=10,367), migraine (N=6,687), dementia (5,675), epilepsy (N=4,558), paralytic ileus and intestinal obstruction (N=2,999), and cardiomyopathy (N=2,342). FinnGen effect estimates and standard errors were used in subsequent Mendelian randomization analyses to define the effect of selected genetic instruments on disease risk.

*Genetic Instrument Selection*

First, genome-wide significant variants from the present European GWAS meta-analysis of mtDNA-CN were chosen (N=383,476). Second, we matched these variants to the

FinnGen v4 GWAS datasets (Feng *et al.*, 2020). Third, to enrich for variants that directly act through mitochondrial processes, we only retained those within 100kb of genes encoding for proteins that are expressed in mitochondria based on MitoCarta3 annotations (Rath *et al.*, 2021). Fourth, we performed LD-pruning in PLINK with 1000Genomes Europeans as the reference panel to ascertain an independent set of genetic variants (LD r2 < 0.01) (Purcell *et al.*, 2007; Abecasis *et al.*, 2012). Lastly, to mitigate potential for horizontal pleiotropy, we further removed variants with strong evidence of acting through alternative pathways by performing a phenome-wide search across published GWAS with Phenoscanner V2 (Kamat *et al.*, 2019). Variants strongly associated with other phenotypes ($P<5x10^{-20}$) were removed unless the variant was a coding mutation located within a gene encoding for the mitochondrial proteome (MitoCarta3) or had an established mitochondrial role based on manual literature review (Rath *et al.*, 2021). A total of 27 genetic variants were used to approximate genetically determined mtDNA-CN levels.

*Mendelian Randomization & Sensitivity Analyses*

Two sample Mendelian Randomization analyses were performed using the "TwoSampleMR" and "MRPRESSO" R packages (Hemani *et al.*, 2018; Verbanck *et al.*, 2018). Effect estimates and standard errors corresponding to the 27 genetic variants on mtDNA-CN (exposure) and mitochondrial disease phenotypes (outcome) were derived from the European GWAS meta-analysis and FinnGen v4 GWAS summary statistics, respectively (S2. Table 9). Three MR methodologies were employed including Inverse Variance Weighted (primary method), Weighted Median, and MR-EGGER methods. MR-PRESSO was used to detect global heterogeneity and P-values were derived based on 1000

simulations. If significant global heterogeneity was detected (P<0.05), a local outlier test was conducted to detect outlying SNPs. After removal of outlying SNPs, MR analyses were repeated. In the absence of heterogeneity (Egger-intercept P $\geq$ 0.05; MR-PRESSO global heterogeneity P $\geq$ 0.05), we reported the inverse-variance weighted result. In the presence of balanced pleiotropy (MR-PRESSO global heterogeneity P < 0.05) and absence of directional pleiotropy (Egger-intercept P $\geq$ 0.05), we reported the weighted median result. In the presence of directional pleiotropy (Egger-intercept P < 0.05), we reported the MR-EGGER result.

**Results**

**AutoMitoC: A streamlined method for array-based mtDNA-CN estimation**

We built on an existing framework for processing normalized SNP probe intensities (L2R values) from genetic arrays into mtDNA-CN estimates known as the "Mitopipeline" (Lane, 2014). The MitoPipeline yields mtDNA-CN estimates that correlate with direct qPCR measurements and has been successfully implemented in several epidemiological investigations(Ashar *et al.*, 2017; Zhang *et al.*, 2017). We developed a novel method, "AutoMitoC", which incorporates three amendments to facilitate large-scale investigations of mtDNA-CN. Firstly, AutoMitoC replaces autosomal signal normalization of common variants with globally rare variants which negates the need for linkage disequilibrium pruning. As a result, this simplifies derivation of mtDNA-CN estimates in ethnically diverse cohorts by allowing for use of a single, universal variant set for normalization. Secondly, to detect potentially cross-hybridizing probes, we empirically assess the association of corrected probe signal intensities with off-target genome intensities, rather

than relying on sequence homology of probe sequences, which is not always available. Lastly, the primary estimate of MT signal is ascertained using principal component analysis (as opposed to using the median signal intensity of MT probes as per the Mitopipeline) which improves concordance of array-based mtDNA-CN estimates with those derived from alternative methods. A detailed description of the AutoMitoC pipeline is provided in Supplementary Results 1.

To benchmark performance of AutoMitoC, array-based mtDNA-CN estimates were compared to complementary measures of mtDNA-CN in two independent studies. Firstly, array-based mtDNA-CN estimates were derived in a subset of 34,436 UKBiobank participants with available whole exome sequencing (WES) data. Reference mtDNA-CN estimates were derived from the proportion of WES reads aligned to the mitochondrial genome relative to the autosome (Longchamps, 2019). AutoMitoC estimates were significantly correlated with WES estimates ($r=0.45$; $P<2.23x10^{-308}$). Since WES data involves enrichment for nuclear coding genes and therefore could result in biased reference estimates for mtDNA-CN, we also performed an independent validation in an ethnically diverse study of 5,791 participants where mtDNA-CN was measured using quantitative PCR, the current gold standard assay (Fazzini *et al.*, 2018). Indeed, we observed stronger correlation between AutoMitoC and qPCR-based estimates ($r=0.64$; $P<2.23x10^{-308}$). Furthermore, AutoMitoC demonstrated robust performance ($r \geq 0.53$) across all ethnic strata in the secondary validation cohort including Europeans (N=2,431), Latin Americans (N=1,704), Africans (N=542), South East Asians (N=471), South Asians (N=186), and others (N=360; S1. Figure 4).

**Genome-wide association study (GWAS) identifies 72 common loci for mtDNA-CN**

A GWAS was performed testing the association between 11,453,766 common genetic variants (MAF>0.005) with mtDNA-CN in 383,476 UKBiobank participants of European ancestry. In total, 9,602 variants were associated with mtDNA-CN at genome-wide significance (Figure 1A; S2. Figure 1), encompassing 82 independent signals in 72 loci (S2. Table 1; S2. Figure 2). The genomic inflation factor was 1.16 and the LD-score intercept was 1.036, indicating that most inflation in test statistics was attributable to polygenicity. Sensitivity analyses revealed that NUMT interference may have played a role in 2 independent signals (2 loci), which were subsequently discarded, leading to a total of 80 independent signals in 70 loci.

Fine-mapping via the FINEMAP algorithm (Benner *et al.*, 2016) yielded 95% credible sets containing 2,363 genome-wide significant variants. Of the 80 independent genetic associations, 17 (22%) mapped to a single candidate causal variant; 32 (39%) mapped to 5 or fewer variants, and 42 (51%) mapped to 10 or fewer variants (Figure 1B; S2. Table 2). Credible sets for 11 genetic signals overlapped with genes responsible for rare mtDNA depletion disorders including *DGUOK* (3), *MGME1* (2), *TFAM* (2), *TWNK* (2), *POLG2* (1), and *TYMP* (1) (S2. Tables 3 & 4). Several associations mapped to coding variants with high posterior probability. *DGUOK* associations mapped to a synonymous variant (rs62641680; Posterior Probability=1) and a non-synonymous variant (rs74874677; PP=1). *TFAM* associations mapped to a 5'UTR variant (rs12247015; PP=1) falling within an ENCODE candidate cis-regulatory element with a promotor-like signature and an intronic variant (rs4397793; PP=1) with a proximal enhancer-like signature. Lastly,

*POLG2* associations mapped to a nonsynonymous variant (rs17850455; PP=1). Beyond the six aforementioned mtDNA depletion genes identified at genome-wide significance, suggestive associations were found for *POLG* (rs2307441; $P=1.0x10^{-7}$), *OPA1* (rs9872432; $P=5.2x10^{-7}$), *SLC25A10* (rs62077224; $P=1.2x10^{-7}$), and *RRM2B* (rs3907099; $P=4.7x10^{-6}$). Given these observations, we hypothesized that mtDNA depletion genes may be generally enriched for common variant associations. Indeed, 10 (53%) of 19 known mtDNA depletion genes (Oyston, 1998) harboured at least suggestive mtDNA-CN associations ($P<5x10^{-6}$).

Additionally, trans-ethnic meta-analysis inclusive of non-Europeans (N=395,781) was performed but given the small increase in sample size, GWAS findings remained highly similar (S2. Figure 3). However, European effect estimates were significantly and highly correlated with those derived from South Asian ($r=0.97$; $P= 2.2x10^{-15}$) and African ($r=0.88$; $P=9.1x10^{-5}$) GWAS analyses (S2. Figure 4).

**mtDNA-CN loci influence mitochondrial gene expression and heteroplasmy**

We postulated that mtDNA-CN loci may regulate copy number by inducing changes in expression of genes that are directly transcribed from mtDNA. Ali *et al*. (2019) recently conducted a GWAS to identify nuclear genetic variants associated with variation in mtDNA-encoded gene expression (i.e. mt-eQTLs) (Ali *et al.*, 2019). Nonsynonymous variants in *LONP1* (rs11085147) and *TBRG4* (rs2304693), as well as an intronic variant in *MRPS35* (rs1127787), were associated with changes in MT gene expression across various tissues (S2. Table 3). Nominally associated mtDNA-CN loci were also observed to influence MT gene expression including intronic variants in both *PNPT1* (rs62165226;

mtDNA-CN P=5.5x10$^{-5}$) and *LRPPRC* (rs10205130; mtDNA-CN P=1.1x10$^{-4}$). Although differences in mitochondrial gene expression may be a consequence rather than a cause of variable mtDNA-CN, the analysis performed by Ali *et al* (2019) was corrected for factors associated with global changes in the mitochondrial transcriptome (Ali *et al.*, 2019). Moreover, the direction of effect estimates between mtDNA-CN and mt-eQTLs varied depending on gene and tissue context. Altogether, such findings imply that some mtDNA-CN loci may regulate mtDNA-CN by influencing mitochondrial gene expression.

A recent GWAS by Nandakumar *et al.* (2021) also reported 20 loci for mtDNA heteroplasmy. While full genome-wide summary statistics were not publicly available to systematically lookup potential effects of the 80 mtDNA-CN GWS variants on mtDNA heteroplasmy, we performed the reverse lookup of whether mtDNA heteroplasmy loci influenced mtDNA-CN. Of 19 matching variants between the GWAS, four heteroplasmy loci were also associated with mtDNA-CN at genome-wide significance including variants nearby or within *TINCR/LONP1* (rs12461806; mtDNA-CN GWAS P=7.5x10$^{-88}$), *TWNK/MPRL43* (rs58678340; P=1.3x10$^{-39}$), *TFAM* (rs1049432; P=1.5x10$^{-21}$), and *PRKAB1* (rs11064881; P=2.6x10$^{-10}$) genes. Consistent with the finding from Nandakumar *et al.* (2021) that the heteroplasmy-increasing *TFAM* allele was also associated with higher mtDNA-CN, we also observed concordant directionality for the other three variants. No additional mtDNA heteroplasmy loci were identified to influence mtDNA-CN when using the suggestive significance threshold.

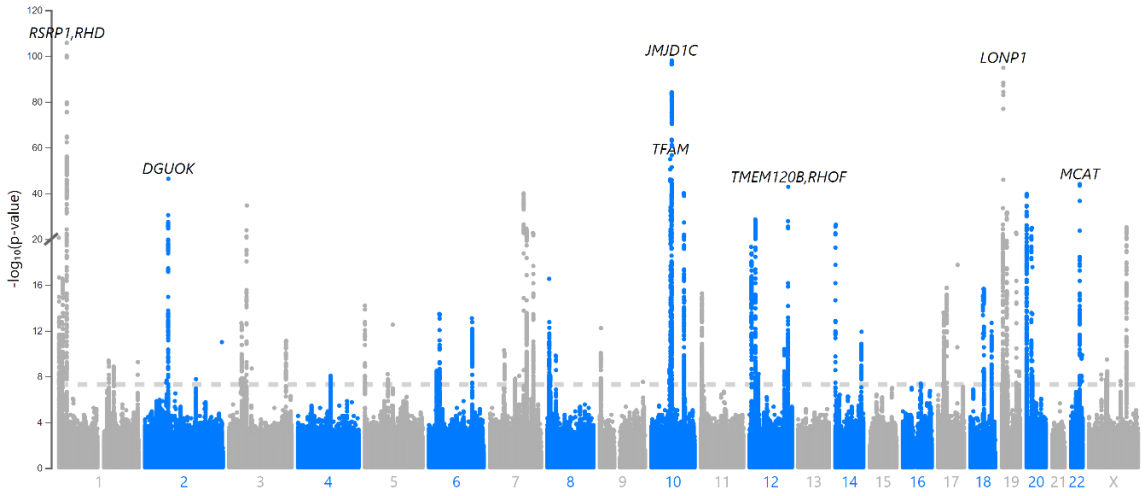
**Genes and pathways implicated in the regulation of mtDNA-CN**

DEPICT analysis led to the prioritization of 91 out of 18,922 genes (FDR $P < 0.05$; S2. Table 4). 87 of these genes intersected with the GeneMANIA database and were uploaded to the GeneMANIA platform to identify additional functionally related genes (Warde-Farley *et al.*, 2010). GeneMANIA analysis discovered an additional 20 related genes (S2. Table 5). Among the 107 total genes prioritized by DEPICT or GeneMANIA (Figure 1C), mitochondrial functions were significantly enriched in gene ontology terms including mitochondrion organization (coverage: 12/225 genes; FDR $P = 7.4 \times 10^{-5}$), mitochondrial nucleoid (6/34; FDR $P = 2.2 \times 10^{-4}$), mitochondrial genome maintenance (4/10; FDR $P = 6.8 \times 10^{-4}$), and mitochondrial matrix (11/257; FDR $P = 6.8 \times 10^{-4}$). Visual inspection of the links between key genes involved in these functions highlights *PPRC1*, a member of the PGC-1A family of mitochondrial biogenesis activators (Richard C. Scarpulla, 2011), as a potential coordinator of mtDNA-related processes (Figure 1C).

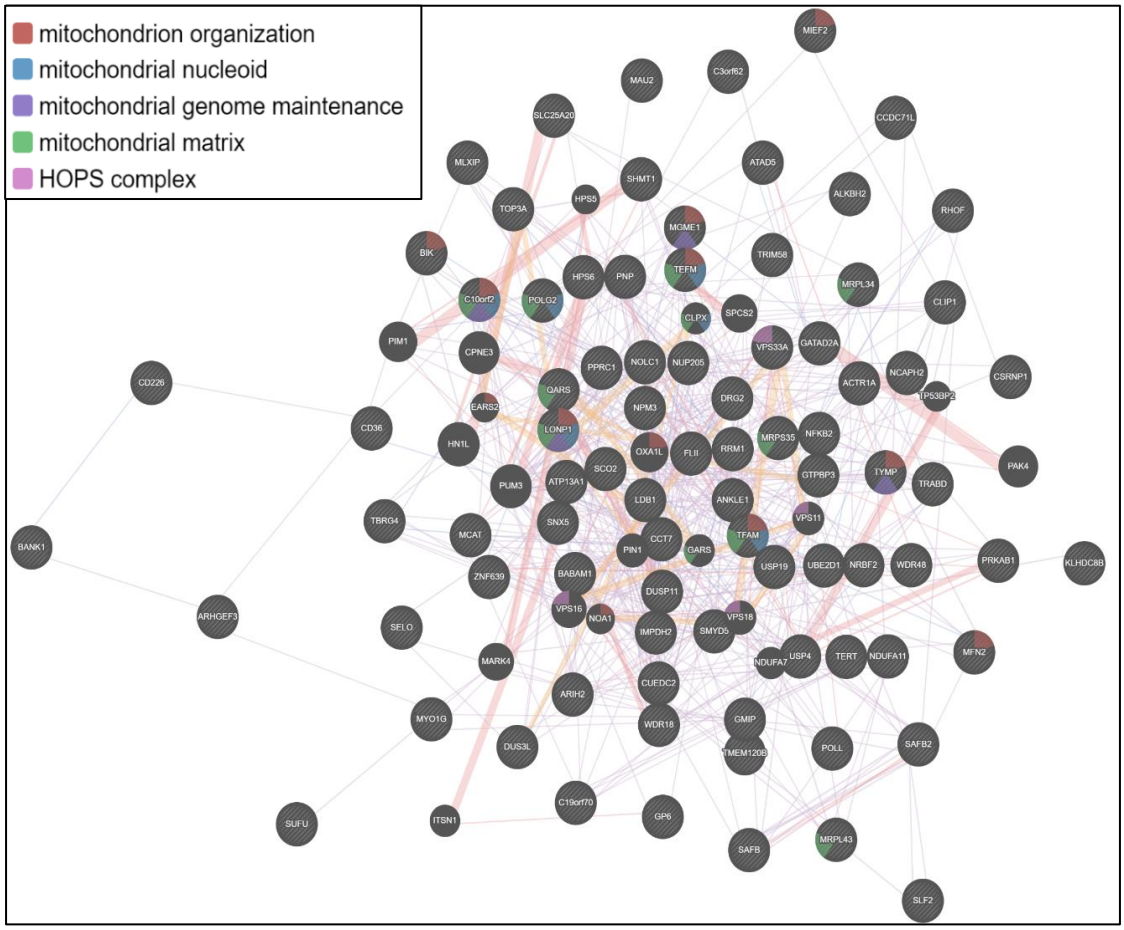
MitoCarta3 is a comprehensive and curated inventory of 1,136 human proteins (1,120 nuclear) known to localize to the mitochondria based on experiments of isolated mitochondria from 14 non-blood tissues (Rath *et al.*, 2021). We leveraged this recently updated database, that was absent from GeneMANIA, to conduct a complementary set of targeted analyses focused on mitochondrial annotations (S2. Table 5). First, we hypothesized that prioritized genes would be generally enriched for genes encoding the mitochondrial proteome. Overall, 27 (25%) of 107 genes had evidence of mitochondrial localization corresponding to a 4.2-fold enrichment (null expectation = 5.9%; $P = 1.0 \times 10^{-10}$). Next, given that *PPRC1*, an activator of mitochondrial biogenesis, was prioritized by DEPICT analyses and then linked to central mtDNA regulators in GeneMANIA, we

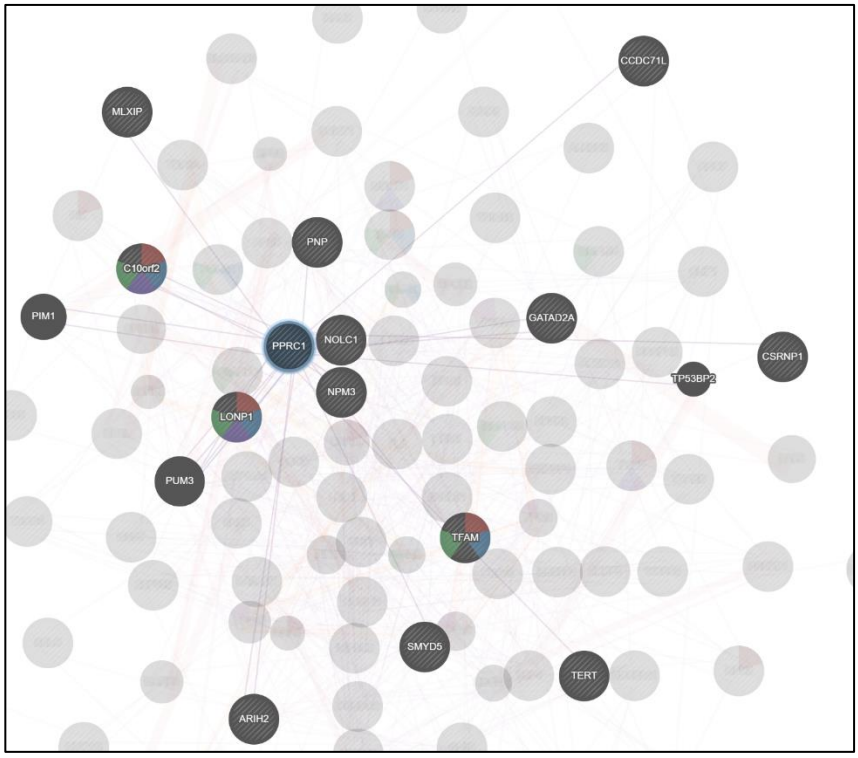
postulated that prioritized genes may be enriched for downstream targets of PGC-1A. PGC-1A induction resulted in a higher mean fold-change among prioritized genes (beta=1.48; 95% CI, 0.60 to 2.37) as compared to any mitochondrial gene (beta=1.19; 95% CI, -0.76 to 3.13; t-test P=0.04). Finally, we categorized the 27 MitoCarta3 genes into their respective pathways. Most (16; 57%) genes were members of the "Mitochondrial central dogma" pathway but other implicated pathways included "Metabolism", "Mitochondrial dynamics and surveillance", "Oxidative phosphorylation", and "Protein import, sorting and homeostasis" pathways (Figure 1D). Four proteins were annotated as part of multiple pathways including *TYMP/SCO2*, *GTPBP3*, *MIEF1*, and *OXA1L* (S2. Table 5).

**A**

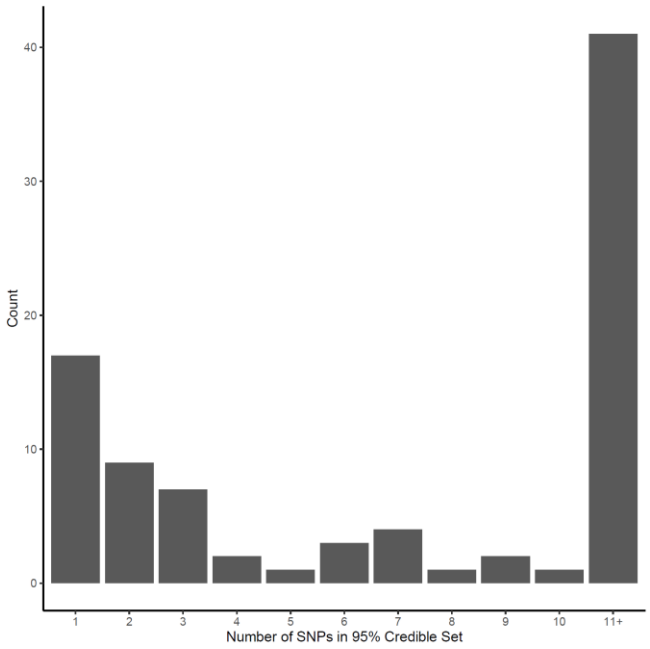


**B**


mitochondrion organization
mitochondrial nucleoid
mitochondrial genome maintenance
mitochondrial matrix
HOPS complex
MIEF2
MAU2
C3orf62
SLC25A20
CCDC71L
MLXIP
ATAD5
SHMT1
ALKBH2
TOP3A
HPS5
RHOF
MGME1
BIK
TRIM58
HPS6
PNP
TEFM
MRPL34
C10orf2
POLG2
CLIP1
CLPX
SPCS2
PIM1
CPNE3
VPS33A
GATAD2A
PPRC1
NOLC1
NUP205
NCAPH2
CSRNP1
CD226
ACTR1A
QARS
TP53BP2
EARS2
NPM3
DRG2
CD36
LONP1
MRPS35
NFKB2
HN1L
OXA1L
FUI
RRM1
PAK4
SCO2
GTPBP3
TYMP
TRABD
PUM3
ATP13A1
LDB1
ANKLE1
TBRG4
SNX5
VPS11
MCAT
CCT7
TFAM
BANK1
PIN1
GARS
ZNF639
BABAM1
USP19
UBE2D1
NRBF2
WDR48
PRKAB1
KLHDC8B
DUSP11
ARHGEF3
VPS16
NOA1
VPS18
SELO
IMPDH2
SMYD5
MARK4
INDUFA7
USP4
TERT
NDUFA11
MFN2
CUEDC2
ARIH2
MYO1G
WDR18
GMIP
DUS3L
TMEM120B
POLL
SAFB2
C19orf70
SUFU
ITSN1
GP6
SAFB
MRPL43
SLF2


CCDC71L
MLXIP
PNP
C10orf2
PIM1
PPRC1
NOLC1
GATAD2A
CSRNP1
TP53BP2
NPM3
LONP1
PUM3
TFAM
SMYD5
TERT
ARIH2

**C**


Count
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11+
Number of SNPs in 95% Credible Set

**D**

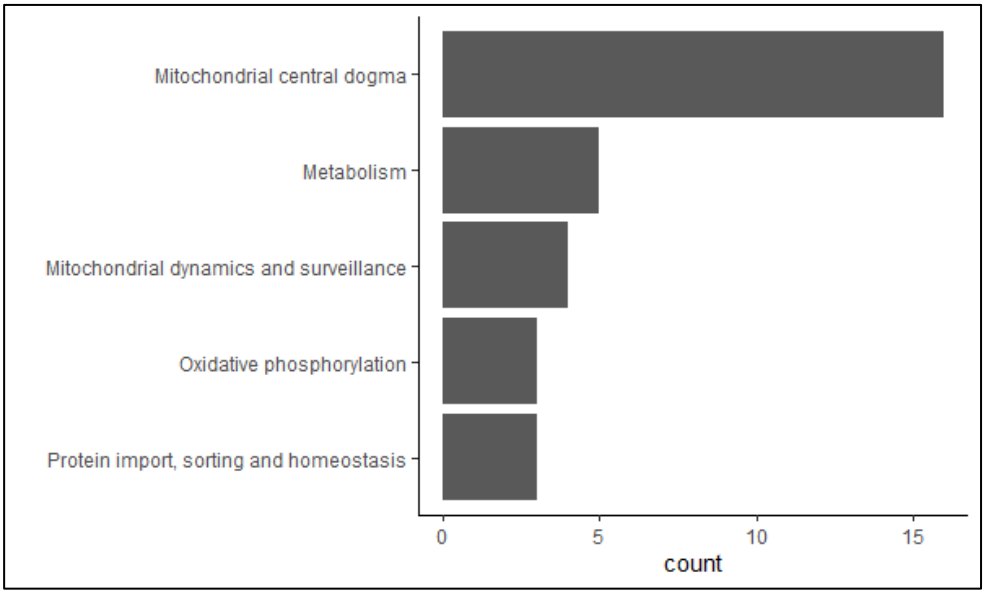



**Figure 1.** Analyses of common genetic loci associated with mtDNA-CN. (A) Manhattan plot illustrating common genetic variant associations with mtDNA-CN. (B) Size distribution of 95% credible sets defined for 80 independent genetic signals. (C) GENE-MANIA-mania protein network interaction exploration (D) "MitoPathway" counts corresponding to 27 prioritized MitoCarta3 genes encoding proteins with known mitochondrial localization.

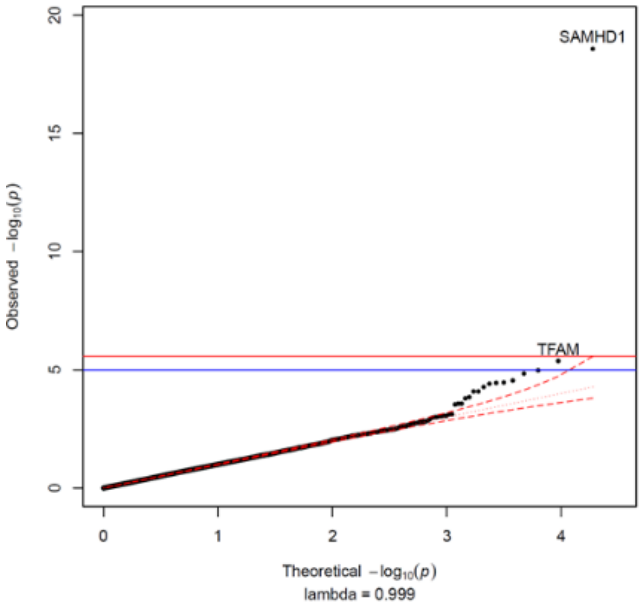
**Exome-wide association testing uncovers rare coding *SAMHD1* mutations as a determinant of mtDNA-CN levels and breast cancer risk**

We performed an exome-wide association study (EXWAS) in 147,740 UKBiobank participants with WES data to assess the contribution of rare coding variants. Among 18,890 genes tested, *SAMHD1* was the only gene reaching exome-wide significance (Figure 2A; S2 Table 8). The carrier prevalence of rare *SAMHD1* mutations was 0.75%,

and on average, mutation carriers had higher mtDNA-CN than non-carriers (beta=0.23 SDs; 95% CI, 0.18-0.29; P=2.6x10$^{-19}$; S2. Figure 5). Also, while none of the 19 known mtDNA depletion genes reached Bonferroni significance, a suggestive association was found for *TFAM* (beta=-0.33; 95% CI, -0.47 to -0.19; P=4.2x10$^{-6}$), and this association was independent of the common *TFAM* variants (rs12247015; rs4397793) previously identified in the GWAS (beta=-0.33; 95% CI, -0.47 to -0.19; P=8.x10$^{-6}$).

To evaluate whether rare *SAMHD1* mutations also influenced disease risk, we conducted phenome-wide association testing of 771 diseases within the UKBiobank. At phenome-wide significance, *SAMHD1* mutation carrier status was associated with approximately two-fold increased risk of breast cancer (OR=1.91; 95% CI, 1.52-2.40; P=2.7x10$^{-8}$), as well as greater risk of "cancer (suspected or other)" (OR=1.52; 95% CI, 1.28-1.80; P=1.1x10$^{-6}$; Figure 2B; S2 Table 9). Exclusion of breast cancer cases attenuated but did not nullify the association with "cancer (suspected or other)" (OR=1.36; 95% CI, 1.10-1.67; P=0.004) suggesting that *SAMHD1* mutations may also increase risk of other cancers, as has been shown for colon cancer (Rentoft *et al.,* 2019). To understand whether differences in mtDNA-CN levels between *SAMHD1* mutation carriers was a consequence of cancer diagnosis, we repeated association testing with mtDNA-CN excluding cancer patients. In this analysis, the association with mtDNA-CN levels was not attenuated (beta=0.26; 95% CI, 0.19-0.32; P=7.8x10$^{-15}$) suggesting that the effect of rare *SAMHD1* variants on mtDNA-CN levels is not driven by its relationship with cancer status. A summary of mitochondrial genes and pathways implicated by common and rare loci is provided in Figure 3.

**A**




**B**

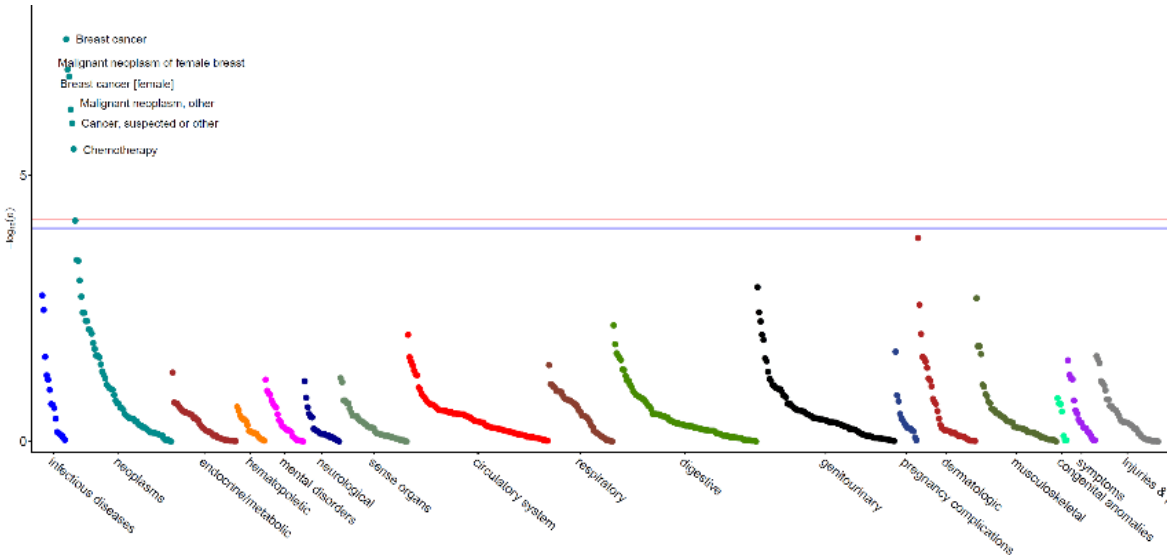



**Figure 2.** Rare variant gene burden association testing with mtDNA-CN and disease risk. (A) QQ plot illustrating expected vs. observed -$\log_{10}$(p) values for exome-wide burden of rare (MAF<0.001) and nonsynymous mutations. (B) Manhattan plot showing phenome-

wide significant associations between *SAMHD1* carrier status and cancer-related phenotypes.

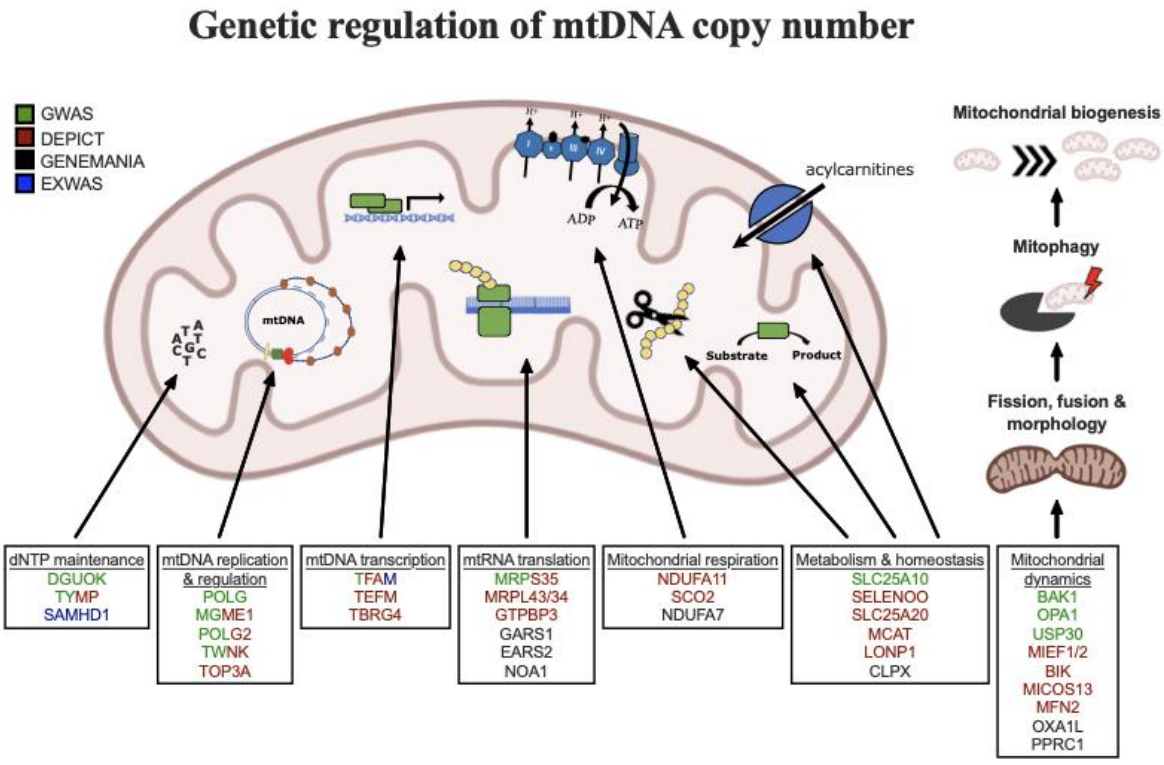



**Figure 3.** Graphical summary of mitochondrial genes and pathways implicated by genetic analyses. Colour-coding indicates through which set(s) of analyses genes were identified. The image was generated using BioRender (https://biorender.com/).

**Mendelian Randomization analysis implicates low mtDNA-CN as a causal mediator of dementia**

Given that common variant loci overlapped with several mtDNA depletion genes, we postulated that polygenically low mtDNA-CN might cause a milder syndrome with phenotypically similar manifestations. To assess whether mtDNA-CN may represent a putative mediator of mtDNA depletion-related phenotypes, we conducted Mendelian

Randomization analyses between genetically determined mtDNA-CN and mitochondrial disease phenotypes using summary statistics derived from the FinnGen v4 GWAS dataset.

After accounting for multiple testing of 10 phenotypes, an association between mtDNA-CN and all-cause dementia was found (OR=1.94 per 1 SD decrease in mtDNA-CN; 95% CI, 1.55-2.32; $P=7.5x10^{-4}$; S2. Table 9; Figure 4). Sensitivity analyses indicated no evidence of global (MR-PRESSO P=0.51; Q-statistic P=0.51) or directional (Egger Intercept P =0.47) pleiotropy. The 27 selected variants accounted for 0.70% of the variance in mtDNA-CN and 0.13% of the risk for dementia, consistent with a causal effect of mtDNA-CN on dementia risk and not vice versa (Steiger $P= 1.9x10^{-62}$). Findings were robust across several different MR methods including the Weighted Median (OR=2.47; 95% CI, 1.93-3.00; P=0.001) and MR-EGGER (OR=2.41; 95% CI, 1.71-3.11; P=0.02) methods. Furthermore, we replicated this result using a second UKBiobank-independent GWAS dataset derived from the International Genomics of Alzheimer's Disease Consortium (2013) including 17,008 Alzheimer's disease patients (OR=1.41; 95% CI, 1.0001-1.98; P=0.04993) (Lambert *et al.*, 2013).

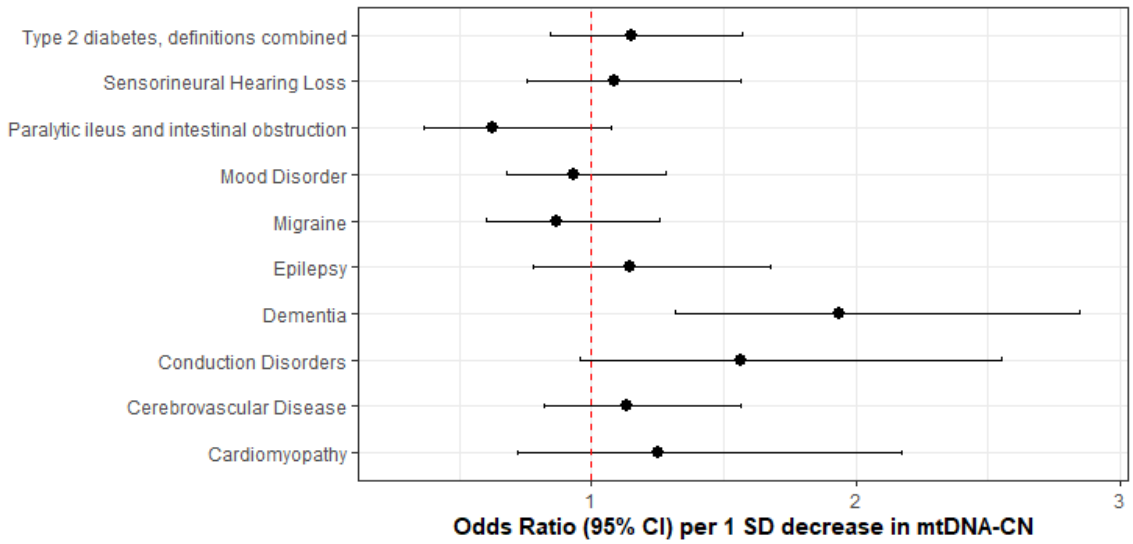



**Figure 4**. Coefficient plots for Mendelian Randomization analyses of mitochondrial disease traits. In the absence of heterogeneity (Egger-intercept P $\geq$ 0.05; MR-PRESSO global heterogeneity P $\geq$ 0.05), the inverse-variance weighted result was reported. In the presence of balanced pleiotropy (MR-PRESSO global heterogeneity P < 0.05), the weighted median result was reported. No set of analyses had evidence for directional pleiotropy (Egger-intercept P < 0.05).

**Discussion**

We developed a novel method to estimate mtDNA-CN from genetic array data, "AutoMitoC", and applied it to the UKBiobank study. Extensive genetic investigations led to several key insights regarding mtDNA-CN. First, several novel common and rare genetic determinants of mtDNA-CN were identified, totalling 71 loci. Second, these loci were enriched for mitochondrial processes related to dNTP metabolism and the replication, packaging, and maintenance of mtDNA. Third, we observed a strong role for common variation within known mtDNA depletion genes in regulating mtDNA-CN in the general

population. Fourth, we found that rare variants in *SAMHD1* not only affect mtDNA-CN levels but also confer risk to cancer. Finally, we provided the first Mendelian Randomization evidence implicating low mtDNA-CN as a causative risk factor for dementia.

While several investigations for mtDNA-CN have been performed, the present study represents the most comprehensive genetic assessment to date(Cai *et al.*, 2015; Longchamps, 2019; Hägg *et al.*, 2020). Notably, Hagg *et al.* (2020) recently conducted a GWAS for mtDNA-CN in 295,150 UKBiobank participants and identified 50 common loci(Hägg *et al.*, 2020). However, the method developed by Hagg *et al.* (2020) calibrated SNP probe intensities based on association with whole-exome sequencing read depths, which may limit the convenience of the method. In contrast, AutoMitoC only necessitates array probe intensities and does not require any secondary genetic measurements (WES or otherwise) for calibration. In addition, AutoMitoC exhibits superior concordance with WES-based estimates (Hagg r=0.33; AutoMitoC r=0.45), which was validated in an independent dataset with gold standard qPCR measurements. Further, Hagg *et al.* (2020) restricted genetic analyses to unrelated European individuals, whereas we incorporated ~100,000 additional individuals and demonstrated consistency in genetic effects between Europeans and non-Europeans ($r \geq 0.88$). The greater sample size in combination with more accurate mtDNA-CN estimates may explain the 44% increase in identified common loci (72 vs 50). Finally, in the present study we included exploration of the role of rare variants through ExWAS and, notably, Mendelian Randomization analyses to assess disease

contexts whereby mtDNA-CN may represent a causal mediator and a potential therapeutic target.

MtDNA-CN has proven to be a biomarker of cardiovascular disease in several large epidemiological studies, with studies often assuming that such relationships are attributable to pathological processes including mitochondrial dysfunction, oxidative stress and inflammation (Tin *et al.*, 2016; Wu *et al.*, 2017; Fazzini *et al.*, 2019; Koller *et al.*, 2020). Consistent with previous GWAS findings, our genetic analyses confirm that mtDNA-CN indeed reflects specific mitochondrial functions, but perhaps not the ones commonly attributable to mtDNA-CN(Cai *et al.*, 2015; Guyatt *et al.*, 2019; Hägg *et al.*, 2020). Primarily, differences in mtDNA-CN reflect mitochondrial processes related to dNTP metabolism and the replication, maintenance, and organization of mtDNA. Secondarily, genes involved in mitochondrial biogenesis, metabolism, oxidative phosphorylation, and protein homeostasis were also identified but do not represent the main constituents. The observed enrichment in common variant associations within mtDNA depletion genes further reinforces the notion that differences in mtDNA-CN first and foremost reflect perturbations in mtDNA-related processes.

No therapy for mtDNA depletion disorders currently exists with treatment mainly consisting of supportive care. Intriguingly, we found that rare variants within *SAMHD1* were associated with increased levels of mtDNA-CN. *SAMHD1* is a multifaceted enzyme with various functions including tumour suppression through DNA repair activity and maintenance of steady-state intracellular dNTP levels, which has been involved in HIV-1 replication (Baldauf *et al.*, 2012; Kretschmer *et al.*, 2015). Rare homozygous and

compound heterozygous loss-of-function mutations in *SAMHD1* result in an immune encephalopathy known as Aicardi Goutiere's syndrome(White *et al.*, 2017). Imbalanced intracellular dNTP pools and chronic DNA damage cause persistent elevations in interferon alpha thus mimicking a prolonged response to HIV-1 infection. While Aicardi Goutiere's syndrome is a severe recessive genetic disorder, case reports of *SAMHD1*-related disease often describe heterozygous parents and siblings as being unaffected or with milder disease (familial chilblain lupus 2) (Haskell *et al.*, 2018). In the UKBiobank, the vast majority (99.4%) of individuals possessing *SAMHD1* mutations were heterozygote carriers, who had a two-fold increased risk of breast cancer. Indeed, our finding that *SAMHD1* mutations associate with both elevated mtDNA-CN levels and risk of breast cancer belies the prevailing notion that higher mtDNA-CN is always a protective signature of proper mitochondrial function and healthy cells. Such findings may have important clinical implications for genetic screening. Firstly, heterozygous SAMHD1 mutations may be an overlooked risk factor for breast cancer considering that ~1 in 130 UKBiobank participants possessed a genetic mutation conferring two-fold elevated risk. Notably, while *SAMHD1* mutations have been described previously to be associated with various cancers (Kohnken, Kodigepalli and Wu, 2015; Rentoft *et al.*, 2019), this gene is not routinely screened nor part of targeted gene panels outside the context of neurological disorders (https://www.genedx.com/test-catalog/available-tests/comprehensive-common-cancer-panel/). Secondly, unaffected parents and siblings of Aicardi Goutiere patients might also present with greater risk of cancer. Thirdly, while SAMHD1 is a highly pleiotropic protein, therapeutic strategies to dampen (but not abolish) SAMHD1 activity might be considered

to treat mtDNA depletion disorders caused by defects in nucleotide metabolism. Indeed, Franzolin *et al.* (2015) demonstrated that siRNA knockdown of SAMHD1 in human fibroblasts with *DGUOK* mtDNA depletion mutations partially recovered mtDNA-CN (Franzolin *et al.*, 2015).

To our knowledge, we provide the first Mendelian Randomization evidence that mtDNA-CN may be causally related to risk of dementia. Although dysfunctional mitochondria have long been implicated in the pathogenesis of Alzheimer's disease, only recently has mtDNA-CN been tested as a biomarker. Silzer *et al.* (2019) conducted a matched case-control study of 46 participants and showed that individuals with cognitive impairment had significantly lower blood-based mtDNA-CN(Silzer *et al.*, 2019). Andrews *et al.* (2020) studied the relationship between post-mortem brain tissue mtDNA-CN and measures of cognitive impairment in 1,025 samples (Andrews and Goate, 2020). Consistent with our findings, a 1 SD decrease in brain mtDNA-CN was associated with lower mini mental state exam (beta = -4.02; 95% CI, -5.49 to -2.55; $P=1.07x10^{-7}$) and higher clinical dementia rating (beta = 0.71; 95% CI, 0.51 to 0.91). Both studies implicate blood and brain-based mtDNA-CN as a marker of dementia but were retrospective. In contrast, Yang *et al.* (2020) observed a significant association between mtDNA-CN and incident risk of neurodegenerative disease (Parkinson's and Alzheimer's disease)(Yang *et al.*, 2021). Altogether, our results combined with previous findings suggest that mtDNA-CN represents both a marker and mediator of dementia. Considering that our overall findings suggest that mtDNA-CN reflects numerous mitochondrial subprocesses, future studies will

be required to disentangle which ones, as reflected by diminished mtDNA-CN, truly mediate dementia pathogenesis.

Several limitations should be noted. First, mtDNA-CN approximated by array-based methods remain imperfectly accurate as compared to qPCR or whole genome sequencing measurements, though we found strong correlation between AutoMitoC and qPCR-based estimates in this study ($r=0.64$; $P<2.23x10^{-308}$). Although whole genome sequencing will eventually supplant array-based mtDNA-CN GWAS, we hypothesize that the improvements made in the areas of speed, portability to ethnically diverse studies, and ease-of-implementation, should greatly increase accessibility of mtDNA-CN research as a plethora of genotyping array data is presently available to re-analyze. Third, Mendelian Randomization analyses were underpowered to conduct a broad survey of diseases in which mitochondrial dysfunction may play a causal role, and equally as important, we were unable to differentiate whether specific mitochondrial subpathways mediated risk of disease. As additional loci are uncovered, such analyses may be feasible. Fourth, variants and genes implicated in the regulation of mtDNA-CN may be specific to blood samples though findings suggest that many mtDNA-CN loci act through genes that are widely expressed in mitochondria across multiple tissues. Future studies are required to determine whether associations are ubiquitous across mitochondria-containing cells and to investigate the role of mtDNA-CN in other tissues. Lastly, whole blood mtDNA-CN reflects a heterogenous mixture of nucleated and unnucleated cells, and despite adjustment for major known confounding cell types, inter-individual differences in cell subpopulations not captured by a standard blood cell count may represent an important source of confounding.

## Conclusion

Although commonly viewed as a simple surrogate marker for the number of mitochondria present within a sample, genetic analyses suggest that mtDNA-CN is a highly complex biomarker under substantial nuclear genetic regulation. mtDNA-CN reflects a mixture of mitochondrial processes mostly pertaining to mtDNA regulation. Accordingly, the true relationship between mtDNA-CN measured in blood samples with human disease remains to be completely defined though we find evidence for mtDNA-CN as a putative causal risk factor for dementia. Future studies are necessary to decipher if mtDNA-CN is directly involved in the pathogenesis of dementia and other diseases or whether other specific mitochondrial processes are truly causative.

## References and Notes:

Abecasis, G. R. *et al.* (2012) 'An integrated map of genetic variation from 1,092 human genomes.', *Nature*, 491(7422), pp. 56–65. doi: 10.1038/nature11632.

Ali, A. T. *et al.* (2019) 'Nuclear genetic regulation of the human mitochondrial transcriptome', *eLife*, 8, pp. 1–23. doi: 10.7554/eLife.41927.

Andrews, S. J. and Goate, A. M. (2020) 'Mitochondrial DNA copy number is associated with cognitive impairment', *Alzheimer's & Dementia*, 16(S5). doi: 10.1002/alz.047543.

Ashar, F. N. *et al.* (2017) 'Association of Mitochondrial DNA Copy Number With Cardiovascular Disease', 21205(11), pp. 1247–1255. doi: 10.1001/jamacardio.2017.3683.

Baldauf, H.-M. *et al.* (2012) 'SAMHD1 restricts HIV-1 infection in resting CD4+ T cells', *Nature Medicine*, 18(11), pp. 1682–1688. doi: 10.1038/nm.2964.

Benner, C. *et al.* (2016) 'FINEMAP: Efficient variable selection using summary data

from genome-wide association studies', *Bioinformatics*, 32(10), pp. 1493–1501. doi: 10.1093/bioinformatics/btw018.

Benner, C. *et al.* (2017) 'Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies', *The American Journal of Human Genetics*, 101(4), pp. 539–551. doi: 10.1016/j.ajhg.2017.08.012.

Burbulla, L. F. *et al.* (2017) 'Dopamine oxidation mediates mitochondrial and lysosomal dysfunction in Parkinson's disease', *Science*, 357(6357), pp. 1255–1261. doi: 10.1126/science.aam9080.

Cai, N. *et al.* (2015) 'Genetic Control over mtDNA and Its Relationship to Major Depressive Disorder', *Current Biology*. The Authors, 25(24), pp. 3170–3177. doi: 10.1016/j.cub.2015.10.065.

Denny, J. C. *et al.* (2013) 'Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data.', *Nature biotechnology*, 31(12), pp. 1102–10. doi: 10.1038/nbt.2749.

Desdín-micó, G. *et al.* (2020) 'T cells with dysfunctional mitochondria induce multimorbidity and premature senescence', 1376(June), pp. 1371–1376.

Fazzini, F. *et al.* (2018) 'Plasmid-normalized quantification of relative mitochondrial DNA copy number', (May), pp. 1–11. doi: 10.1038/s41598-018-33684-5.

Fazzini, F. *et al.* (2019) 'Mitochondrial DNA copy number is associated with mortality and infections in a large cohort of patients with chronic kidney disease', 8, pp. 480–488. doi: 10.1016/j.kint.2019.04.021.

Feng, Y.-C. A. *et al.* (2020) 'Findings and insights from the genetic investigation of age

of first reported occurrence for complex disorders in the UK Biobank and FinnGen', *medRxiv*, p. 2020.11.20.20234302. Available at: https://doi.org/10.1101/2020.11.20.20234302.

Franzolin, E. *et al.* (2015) 'The deoxynucleoside triphosphate triphosphohydrolase activity of SAMHD1 protein contributes to the mitochondrial DNA depletion associated with genetic deficiency of deoxyguanosine kinase', *Journal of Biological Chemistry*, 290(43), pp. 25986–25996. doi: 10.1074/jbc.M115.675082.

Gene, T. and Consortium, O. (2000) 'Gene Ontology : tool for the', 25(may), pp. 25–29.

Gorman, G. S. *et al.* (2016) 'Mitochondrial diseases', *Nature Reviews Disease Primers*. Macmillan Publishers Limited, 2. doi: 10.1038/nrdp.2016.80.

Guyatt, A. L. *et al.* (2019) 'A genome-wide association study of mitochondrial DNA copy number in two population-based cohorts'. Human Genomics, pp. 1–17.

Hägg, S. *et al.* (2020) 'Deciphering the genetic and epidemiological landscape of mitochondrial DNA abundance', *Human Genetics*. Springer Berlin Heidelberg, (0123456789). doi: 10.1007/s00439-020-02249-w.

Haskell, G. T. *et al.* (2018) 'Combination of exome sequencing and immune testing confirms Aicardi–Goutières syndrome type 5 in a challenging pediatric neurology case', *Molecular Case Studies*, 4(5), p. a002758. doi: 10.1101/mcs.a002758.

Hemani, G. *et al.* (2018) 'The MR-base platform supports systematic causal inference across the human phenome', *eLife*, 7, pp. 1–29. doi: 10.7554/eLife.34408.

Hu, L., Yao, X. and Shen, Y. (2016) 'Altered mitochondrial DNA copy number contributes to human cancer risk: Evidence from an updated meta-analysis', *Scientific*

*Reports*. Nature Publishing Group, 6(October), pp. 1–11. doi: 10.1038/srep35859.

Jagadeesh, K. A. *et al.* (2016) 'M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity.', *Nature genetics*, 48(12), pp. 1581–1586. doi: 10.1038/ng.3703.

Kamat, M. A. *et al.* (2019) 'PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations.', *Bioinformatics (Oxford, England)*, 35(22), pp. 4851–4853. doi: 10.1093/bioinformatics/btz469.

Kim, C. *et al.* (2015) 'Mitochondrial DNA copy number and chronic lymphocytic leukemia/small lymphocytic lymphoma risk in two prospective studies', *Cancer Epidemiology Biomarkers and Prevention*, 24(1), pp. 148–153. doi: 10.1158/1055-9965.EPI-14-0753.

Kohnken, R., Kodigepalli, K. M. and Wu, L. (2015) 'Regulation of deoxynucleotide metabolism in cancer: Novel mechanisms and therapeutic implications', *Molecular Cancer*, 14(1), pp. 1–11. doi: 10.1186/s12943-015-0446-6.

Koller, A. *et al.* (2020) 'Mitochondrial DNA copy number is associated with all-cause mortality and cardiovascular events in patients with peripheral arterial disease', *Journal of Internal Medicine*, 287(5), pp. 569–579. doi: 10.1111/joim.13027.

Kretschmer, S. *et al.* (2015) 'SAMHD1 prevents autoimmunity by maintaining genome stability', *Annals of the Rheumatic Diseases*, 74(3). doi: 10.1136/annrheumdis-2013-204845.

Lambert, J. C. *et al.* (2013) 'Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.', *Nature genetics*, 45(12), pp. 1452–8. doi:

10.1038/ng.2802.

Lane, J. (2014) *MitoPipeline: Generating Mitochondrial copy number estimates from SNP array data in Genvisis*. Available at: http://genvisis.org/MitoPipeline/.

Longchamps, R. J. (2019) 'EXPLORING THE ROLE OF MITOCHONDRIAL DNA QUANTITY AND QUALITY', *Biorxiv*, (August).

Longchamps, R. J. *et al.* (2020) 'Evaluation of mitochondrial DNA copy number estimation techniques', *PLOS ONE*. Edited by D. C. Samuels, 15(1), p. e0228166. doi: 10.1371/journal.pone.0228166.

Mbatchou, J. *et al.* (2020) 'Computationally efficient whole genome regression for quantitative and binary traits', pp. 1–88. doi: 10.1101/2020.06.19.162354.

Nandakumar, P. *et al.* (2021) 'Nuclear genome-wide associations with mitochondrial heteroplasmy', *Science Advances*, 7(12), pp. 1–10. doi: 10.1126/sciadv.abe7520.

Oyston, J. (1998) 'Online Mendelian Inheritance in Man.', *Anesthesiology*, 89(3), pp. 811–2.

Pers, T. H. *et al.* (2015) 'Biological interpretation of genome-wide association studies using predicted gene functions', *Nature Communications*, 6, p. 5890. doi: 10.1038/ncomms6890.

Purcell, S. *et al.* (2007) 'PLINK: a tool set for whole-genome association and population-based linkage analyses.', *American journal of human genetics*, 81(3), pp. 559–75. doi: 10.1086/519795.

Rath, S. *et al.* (2021) 'MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations', *Nucleic acids research*. Oxford

University Press, 49(D1), pp. D1541–D1547. doi: 10.1093/nar/gkaa1011.

Rentoft, M. *et al.* (2019) 'Erratum: Heterozygous colon cancer-associated mutations of SAMHD1 have functional significance (Proceedings of the National Academy of Sciences of the United States of America (2016) 113 (4723-4728) DOI: 10.1073/pnas.1519128113)', *Proceedings of the National Academy of Sciences of the United States of America*, 116(10), p. 4744. doi: 10.1073/pnas.1902081116.

Richard C. Scarpulla (2011) 'Metabolic control of mitochondrial biogenesis through the PGC-1 family regulatory network', *Biochim Biophys Acta.*, 1813(7), pp. 1269–1278. doi: 10.1016/j.bbamcr.2010.09.019.

Silzer, T. *et al.* (2019) 'Circulating mitochondrial DNA : New indices of type 2 diabetes-related cognitive impairment in Mexican Americans', pp. 1–21.

Simone, D. *et al.* (2011) 'The reference human nuclear mitochondrial sequences compilation validated and implemented on the UCSC genome browser.', *BMC genomics*, 12, p. 517. doi: 10.1186/1471-2164-12-517.

Sliter, D. A. *et al.* (2020) 'Parkin and PINK1 mitigate STING-induced inflammation', *Nature*, 561(7722), pp. 258–262. doi: 10.1038/s41586-018-0448-9.Parkin.

Sudlow, C. *et al.* (2015) 'UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.', *PLoS medicine*, 12(3), p. e1001779. doi: 10.1371/journal.pmed.1001779.

Tin, A. *et al.* (2016) 'Association between mitochondrial DNA copy number in peripheral blood and incident CKD in the atherosclerosis risk in communities study', *Journal of the American Society of Nephrology*, 27(8), pp. 2467–2473. doi: 10.1681/ASN.2015060661.

Verbanck, M. *et al.* (2018) 'Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases', *Nature Genetics*. Springer US, 50(5), pp. 693–698. doi: 10.1038/s41588-018-0099-7.

Vuckovic, D. *et al.* (2020) 'The Polygenic and Monogenic Basis of Blood Traits and Diseases.', *Cell*, 182(5), pp. 1214-1231.e11. doi: 10.1016/j.cell.2020.08.008.

Warde-Farley, D. *et al.* (2010) 'The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function', *Nucleic Acids Research*, 38(SUPPL. 2), pp. 214–220. doi: 10.1093/nar/gkq537.

Wei, W. Q. *et al.* (2017) 'Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record', *PLoS ONE*, 12(7), pp. 1–16. doi: 10.1371/journal.pone.0175508.

White, T. E. *et al.* (2017) 'A SAMHD1 mutation associated with Aicardi-Goutières syndrome uncouples the ability of SAMHD1 to restrict HIV-1 from its ability to downmodulate type I interferon in humans', *Human Mutation*, 38(6), pp. 658–668. doi: 10.1002/humu.23201.

Willer, C. J., Li, Y. and Abecasis, G. R. (2010) 'METAL: Fast and efficient meta-analysis of genomewide association scans', *Bioinformatics*, 26(17), pp. 2190–2191. doi: 10.1093/bioinformatics/btq340.

Wu, I.-C. *et al.* (2017) 'Interrelations Between Mitochondrial DNA Copy Number and Inflammation in Older Adults', *The Journals of Gerontology: Series A*, 72(7), pp. 937–944. doi: 10.1093/gerona/glx033.

Wu, P. *et al.* (2019) ‘Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation.’, *JMIR medical informatics*, 7(4), p. e14325. doi: 10.2196/14325.

Yang, S. Y. *et al.* (2021) ‘Blood-derived mitochondrial DNA copy number is associated with gene expression across multiple tissues and is predictive for incident neurodegenerative disease’, *Genome Research*. doi: 10.1101/gr.269381.120.

Zhang, Y. *et al.* (2017) ‘Association between mitochondrial DNA copy number and sudden cardiac death : findings from the Atherosclerosis Risk in Communities study ( ARIC )’, pp. 3443–3448. doi: 10.1093/eurheartj/ehx354.

**Acknowledgements:** We want to acknowledge the participants and investigators of the UKBiobank and FinnGen studies.

**Funding:**

Canadian Institutes of Health Research Frederick Banting and Charles Best Canada Graduate Scholarships Doctoral Award (MC, PM)

Canadian Institutes of Health Research Post-Doctoral Fellowship Award (RM)

Wellcome Trust Grant number: 099313/B/12/A; Crasnow Travel Scholarship; Bongani Mayosi UCT-PHRI Scholarship 2019/2020 (TM)

Wellcome Trust Health Research Board Irish Clinical Academic Training (ICAT) Programme Grant Number: 203930/B/16/Z (CJ)

European Research Council COSIP Grant Number: 640580 (MO)

E.J. Moran Campbell Internal Career Research Award (MP)

CISCO Professorship in Integrated Health Systems (GP)

Canada Research Chair in Genetic and Molecular Epidemiology (GP)

**Author Contributions:**

Conceptualization: MC, GP

Methodology: MC, WN, GP

Bioinformatics and Statistical Analysis: MC, PM, NP, GP, SN, IK, RL, MK, TM

Visualization: MC, PM, NP, RM, CJ

Data Interpretation: MC, PM, GP, SN, LA, RM

Funding or data acquisition: GP, MO, NC

Project administration: GP

Supervision: GP

Writing – original draft: MC

Writing – review & editing: MC, PM, NP, WN, RM, SN, IK, RL, MK, CJ, TM, NC, MO, MP, LA, GP

**Competing Interests:** None

**Data and materials availability:** Individual-level UKBiobank genotypes and phenotypes can be acquired upon successful application (https://bbams.ndph.ox.ac.uk/ams/). All individual-level UKBiobank data was accessed as part of application # 15255. FinnGen summary statistics are freely available to download (https://www.finngen.fi/en/access_results). All data products generated as part of this study will be made publicly accessible. Specifically, a software package for the AutoMitoC array-based mtDNA-CN estimation pipeline will be put on GitHub. The mtDNA-CN estimates derived in UKBiobank participants will be returned to the UKBiobank and made accessible

to researchers through the data showcase (https://biobank.ndph.ox.ac.uk/showcase/). Summary-level association statistics from GWAS will be made publicly available for download. All remaining data are available in the main text or supplementary materials.

**Ethics Statement:** Approval was received to use UKBiobank study data in this work under application ID # 15255 ("Identification of the shared biological and sociodemographic factors underlying cardiovascular disease and dementia risk"). The UKBiobank study obtained ethics approval from the North West Multi-centre Research Ethics Committee which encompasses the UK (REC reference: 11/NW/0382). All research participants provided informed consent.

# CHAPTER 5:

## Mitochondrial DNA copy number as a marker and mediator of stroke prognosis: Observational and Mendelian Randomization analyses

Submitted to *JAMA Neurology*. (May 29th, 2021)

**Mitochondrial DNA copy number as a marker and mediator of stroke prognosis: Observational and Mendelian Randomization analyses**

**Authors:** Michael Chong[1,2,3,4], MSc; Sukrit Narula[1,2,5], BA; Robert Morton[1,2,4], PhD; Conor Judge[1,2,7], MB; Loubna Akhabir[1,2,6], PhD; Nathan Cawte[1,2], BSc; Nazia Pathan[1,2], BSc; Ricky Lali[1,2,5], MSc; Pedrum Mohammadi-Shemirani[1,2,4], BSc; Ashkan Shoamanesh, MD[1,6]; Martin O'Donnell[1,7], PhD; Salim Yusuf[1,2,5,6], DPhil; Peter Langhorne[8], PhD; Guillaume Paré[1,2,3,4,5,6]*, MD.

Dr. Guillaume Paré*
Population Health Research Institute
237 Barton St East
Hamilton, ON Canada L8L 2X2
Email: pareg@mcmaster.ca

[1]Population Health Research Institute (PHRI), David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton Health Sciences; 237 Barton Street East, Hamilton, L8L2X2, Ontario, Canada.

[2]Thrombosis and Atherosclerosis Research Institute; 237 Barton Street East, Hamilton, L8L2X2, Ontario, Canada.

[3]Department of Biochemistry and Biomedical Sciences, McMaster University; 1280 Main Street West, Hamilton, Ontario, L8S 4K1, Canada

[4]Department of Pathology and Molecular Medicine, Michael G. DeGroote School of Medicine, McMaster University; 1280 Main Street West, Hamilton, L8S 4K1, Ontario, Canada

[5]Department of Health Research Methods, Evidence, and Impact, McMaster University; 1280 Main Street West, Hamilton, L8S 4K1, Ontario, Canada

[6]Department of Medicine, McMaster University, Michael G. DeGroote School of Medicine; 1280 Main Street West, Hamilton, L8S 4K1, Ontario, Canada

[7]National University of Ireland Galway; University Road, H91 TK33, Galway, Ireland.

[8]Institute of Cardiovascular and Medical Sciences, University of Glasgow; 126 University Place, G1286A, Glasgow, UK

**Manuscript Word Count**: 3797

## Key Points

**Question:** Do stroke patients with lower buffy coat mitochondrial DNA copy number (mtDNA-CN) have worse prognosis?

**Findings**: Stroke patients with lower mtDNA-CN levels measured within one week of stroke onset had significantly higher odds of worse outcomes at 1-month follow-up, including poor functional outcome (modified Rankin Scale [mRS] 3-6) and mortality. Two-sample Mendelian Randomization analyses in independent datasets revealed a significant association between genetic predisposition to lower mtDNA-CN and higher risk of poor functional outcomes at 3-months follow-up.

**Meaning:** Our findings suggest that low mtDNA-CN is a prognostic marker and a putative causal determinant of post-stroke outcomes.


## Abstract

**Importance:** Low buffy coat mitochondrial DNA copy number (mtDNA-CN) is associated with incident risk of stroke and post-stroke mortality; however, its prognostic utility as a marker of post-stroke outcomes has not been extensively explored, nor is it known whether mtDNA-CN is a causal determinant.

**Objective:** To investigate whether low buffy coat mtDNA-CN is a marker and causal determinant of post-stroke outcomes using epidemiological and genetic studies.

**Design and Setting:** This study comprised a two-stage analysis. First, we performed association testing between baseline buffy coat mtDNA-CN measurements and 1-month post-stroke outcomes in 3498 acute, first stroke cases from 25 countries from the international, multicenter case-control study, “Importance of Conventional and Emerging

Risk Factors of Stroke in Different Regions and Ethnic Groups of the World" (INTERSTROKE). Then, we performed two-sample Mendelian Randomization analyses to evaluate potential causative effects of low mtDNA-CN on 3-month stroke outcomes. Genetic variants associated with mtDNA-CN levels were derived from the UKBiobank study (N=383476), and corresponding effects on 3-month stroke outcomes were ascertained from the Genetics of Ischemic Stroke funCtional Outcome study (GISCOME; N=6021).

**Main Outcomes and Measures:** We hypothesized that measured and genetically determined mtDNA-CN are associated with mRS-based outcomes.

**Results:** Independent of baseline stroke severity, a 1- standard deviation (SD) lower mtDNA-CN at baseline was associated with increased odds of greater 1-month disability (ordinal mRS; OR=1.16; 95% CI, 1.08-1.24; $P=4.4x10^{-5}$), poor functional outcome status (mRS 3-6 vs. 0-2; OR=1.21; 95% CI, 1.08-1.34; $P=6.9x10^{-4}$), and mortality (OR=1.35; 95% CI, 1.14-1.59; $P=3.9x10^{-4}$). Subgroup analyses demonstrated consistent effects across stroke type, sex, age, country income level, and education level. In addition, mtDNA-CN significantly improved reclassification of poor functional outcome status (Net Reclassification Index (NRI)=0.16; 95% CI, 0.08-0.23; $P=3.6x10^{-5}$) and mortality (NRI=0.31; 95% CI, 0.19-0.43; $P=1.7x10^{-7}$) beyond known prognosticators. Using independent datasets, Mendelian Randomization revealed that a 1 SD decrease in genetically determined mtDNA-CN was associated with increased odds of greater 3-month disability quantified by ordinal mRS (OR=2.35; 95% CI, 1.13-4.90; P=0.02) and poor functional outcome status (OR=2.68; 95% CI, 1.05-6.86; P=0.04).

**Conclusions and Relevance:** Buffy coat mtDNA-CN is a novel and robust marker of post-stroke prognosis that may also be a causal determinant of post-stroke outcomes.

## Introduction

Stroke patients from low and middle-income countries bear a disproportionate burden of post-stroke complications[1–3]. As such, identifying cost-effective and highly predictive biomarkers that mediate post-stroke recovery will allow for better risk stratification and novel targets for acute stroke treatment[4].

Mitochondrial health has an important role in both stroke pathogenesis and recovery[5,6], and mitochondrial function can be measured using an inexpensive and accessible assay that quantifies the ratio of mitochondrial to nuclear DNA copies, known as mitochondrial DNA copy number (mtDNA-CN). Rare genetic disorders characterized by severe loss of mtDNA-CN, formally referred to as "mtDNA depletion" syndromes, can cause migraine, leukoencephalopathy, and stroke-like episodes[7,8]. In the broader population, perturbations of leukocyte mtDNA-CN have been reported to reflect general mitochondrial dysfunction, oxidative stress, impaired oxidative phosphorylation, and inflammation[9]. Indeed, low leukocyte mtDNA-CN is associated with increased risks of secondary hospitalization and mortality in patients with atherosclerotic and chronic kidney disease[10–12]. To our knowledge, only one study has investigated the association between mtDNA-CN and post-stroke outcomes, wherein a prospective cohort study of 1484 Chinese stroke patients reported an association between mtDNA-CN and mortality[13]. While these findings suggest a potential role for mtDNA-CN as a risk factor for post-stroke outcomes, and in particular mortality, several important questions remain to be addressed regarding:

(i) the robustness of associations across stroke type and other clinically relevant subgroups, (ii) whether associations are independent of baseline stroke severity, (iii) if mtDNA-CN is associated with the degree of functional disability among stroke survivors, and (iv) if mtDNA-CN is a causal determinant of post-stroke outcomes.

To address these questions, we investigated the relationships between both measured and genetically predicted mtDNA-CN levels with post-stroke outcomes using large-scale datasets. First, we evaluated the association between buffy coat mtDNA-CN levels measured within one week of stroke symptom onset and 1-month outcomes in 3498 stroke patients from the "Importance of Conventional and Emerging Risk Factors of Stroke in Different Regions and Ethnic Groups of the World" (INTERSTROKE) study[14]. Second, to assess whether lower mtDNA-CN levels may be a causal risk factor for poor outcomes at 3-months after stroke, we conducted two-sample Mendelian Randomization (MR) analyses using genetic effects derived from the UKBiobank (N=383476)[15] and the Genetics of Ischemic Stroke funCtional Outcome (GISCOME; N=6165)[16]. Overall, we explored whether low mtDNA-CN represents a marker and casual driver of poor post-stroke outcomes.

## Methods

### *INTERSTROKE*

INTERSTROKE is a large international case-control study encompassing 32 countries across Asia, North America, South America, Europe, Australia, and Africa[14]. The study design has been described in detail previously[17]. In brief, participants were enrolled between January 11, 2007 and August 8, 2015. Cases consisted of patients with acute, first

stroke (ischemic or hemorrhagic) presenting within 5 days of symptom onset and 72 hours of hospital admission. Strokes were defined according to the World Health Organization definition, and subtypes were confirmed by neuroimaging (CT or MRI). Demographic characteristics, medical history, and risk factor data were collected through standardized questionnaires and physical examination. For patients who could not communicate, a proxy respondent was used (spouse or first-degree relative living in the same household aware of the patient's medical history and current treatments). All participants (or their proxies) provided written informed consent. The modified-Rankin scale (mRS)[18] was used as a marker of stroke severity and was measured at baseline and at 1-month follow-up. The presence of hemorrhagic transformation after ischemic stroke was assessed through neuroimaging (either CT or MRI) and adjudicated locally by a site investigator. The present analyses were performed on a subset of 3498 INTERSTROKE cases with qPCR mtDNA-CN measurements.

***MtDNA-CN Measurement and Quality Control***

At each recruitment centre, non-fasting peripheral blood samples were collected in EDTA whole blood tubes from stroke patients within one week of symptom onset (and within 72 hours of hospital admission). Blood samples were shipped to the Clinical Research Laboratory and Biobank, located in Hamilton, Ontario, Canada, where DNA extraction was performed. DNA was extracted from the buffy coat layer of centrifuged samples using the QIAGEN QIAsymphony DNA Midi (96.7%), DNA Mini (2.7%) or DSP DNA Midi (0.6%) kits. mtDNA-CN was assayed by the Genetic and Molecular Epidemiology Lab located in Hamilton, Ontario, Canada using a plasmid-normalized

quantitative Polymerase Chain Reaction (qPCR) method developed by Fazzini *et al.* (2018)[19]. Upon visual inspection of the distribution of mtDNA-CN values, a single sample with an extreme outlying value was removed. Additional outliers beyond 3 standard deviations (SD) of the mean were winsorized to the 99.7$^{th}$ percentile. MtDNA-CN values were normalized for known confounders by taking the residuals from a linear regression model for mtDNA-CN (dependent variable) versus age, sex, ethnicity, and qPCR batch (independent variables). The resulting numerical representation of mtDNA-CN was standardized to a mean of 0 and SD of 1 for subsequent analyses.

***Statistical Analysis***

All statistical analyses were performed using the statistical programming language 'R' (version 3.6.0). Plots were generated using a combination of the "ggplot2", "viridis", "dplyr", "grid", and "gridExtra" R packages. In INTERSTROKE, association testing was conducted to assess the relationship between low mtDNA-CN at baseline (continuous variable or discretized into quartiles) and stroke markers at two timepoints: 1) markers collected at the time of the stroke event (hereafter referred to as 'baseline' severity markers) and 2) markers collected 1-month after the stroke event. The primary marker of baseline stroke severity was ordinal mRS. Secondary markers included level of consciousness (alert, drowsy, or unconscious) and hemorrhagic transformation after ischemic stroke. The primary stroke outcome at 1-month follow-up was ordinal mRS. Secondary outcomes at 1-month follow-up included other formulations of mRS, specifically, poor functional outcome status (dichotomized mRS 3-6 vs. 0-2) and mortality status. Ordinal regression was used for analysis between ordinal mRS and consciousness ("polr" R package). The

proportional odds assumption was evaluated using the Brant test ("Brant" R package). Logistic regression analysis was conducted for dichotomous variables including hemorrhagic transformation at baseline and 1-month post-stroke outcomes (poor functional outcome and mortality statuses). All regression models were adjusted for age, sex, region, education level (none or primary school vs. high school, trade school, college, or university), 2018 World Bank country income stratum (high, upper-middle, and lower-middle or low income), household income (adjusted for country), primary stroke type (ischemic vs. hemorrhagic stroke) and ischemic stroke Oxfordshire Community Stroke Project (OCSP) classification, pre-stroke dependency (pre-stroke mRS 3-5 vs. 0-2), Charleson comorbidity index, and stroke risk factors (hypertension, diabetes, hypercholesterolemia, atrial fibrillation or flutter, current smoker status, and waist to hip ratio) as defined previously[14]. In addition to these covariates, baseline stroke severity (baseline mRS) was additionally included in models for 1-month post-stroke outcomes. For analysis of dichotomous outcomes, additional subgroup analyses were performed stratifying by primary stroke type, sex, age ($< 65$ vs. $\geq 65$ years), country income level, and education level. The Net Reclassification Index (NRI) was used to assess model reclassification improvement upon addition of mtDNA-CN to a baseline model including the following covariates: age, sex, region, education level, country income level, household income, primary stroke type and OCSP classification, pre-stroke dependency, Charleson comorbidity index, hypertension, diabetes, hypercholesterolemia, atrial fibrillation or flutter, current smoker status, and waist to hip ratio ("Hmisc" R package).

***Mendelian Randomization***

Mendelian Randomization (MR) is a statistical genetics framework that leverages the random assortment of genetic alleles (Mendel's second law of independent assortment) to perform causal inference between an exposure and an outcome [20–22]. The use of randomized, genetic alleles as instrumental variables for an exposure endows several advantages including robustness to traditional confounding factors and reverse causation. Indeed, evidence from animal models suggests that stroke induces changes in mtDNA-CN levels, and therefore reverse causality is a relevant concern that is addressed by MR[23,24]. To evaluate the potential causal relationship between low mtDNA-CN (exposure) and stroke prognosis (outcome), we performed "two-sample" MR analyses incorporating summary-level GWAS data from two independent studies. Genetic variants associated with mtDNA-CN levels were identified from a previous genome-wide association study (GWAS) we conducted in 383476 Caucasian participants from the UKBiobank study[25]. UKBiobank is a prospective cohort study including UK residents (ages 40-69 years) recruited from 2006-2010[26]. Eligibility criteria for the mtDNA-CN GWAS included Caucasian participants with suitable genetic microarray data who had non-outlying blood cell count and array intensity values[25]. UKBiobank mtDNA-CN estimates were derived using AutoMitoC, a computational pipeline that leverages array-based data to estimate mtDNA-CN levels[25]. Corresponding genetic effects on 3-month mRS were obtained from the Genetics of Ischaemic Stroke Functional Outcome (GISCOME) GWAS. GISCOME included 6021 Caucasian ischemic stroke patients from 12 studies across Europe, the United States, and Australia[16]. Two formulations of 3-month mRS were tested in the present study: ordinal mRS and poor functional outcome status (mRS 3-6 vs. 0-2). In

GISCOME, 2280 (63%) participants suffered poor functional outcome. There is no sample overlap between UKBiobank and GISCOME datasets.

As previously described[25], an independent set of 26 genome-wide significant variants associated with mtDNA-CN located nearby or within genes expressed in the mitochondria were selected as instruments to genetically approximate mtDNA-CN levels (S. Methods; S. Tables 6 & 7). Collectively, these variants had an F-statistic of 100 which is sufficient (F>10) for the purposes of identifying a causal effect.

Two-sample MR analyses were executed using the "TwoSampleMR" (version 0.5.5) and "MRPRESSO" (version 1.0) R packages[21,33]. Three MR methods were employed including the inverse variance weighted, weighted median, and MR-Egger methods. MR-PRESSO was used to detect global heterogeneity with P-values derived based on 1000 simulations. The Egger intercept test was used to assess directional pleiotropy. None of the MR associations exhibited significant heterogeneity (MR-PRESSO Global Test $P > 0.05$) or directional pleiotropy (Egger intercept $P > 0.05$), and thus the causal effect estimates from the Inverse Variance Weighted method was reported for all MR analyses. Causal effects were expressed as odds of a higher mRS category (or of poor functional outcome) per 1 standard deviation decrease in genetically determined mtDNA-CN levels.

Phenotypic mtDNA-CN associations may also, in part, capture differences in blood cell proportions[29,30], so we also examined the relationship between genetically determined blood cell traits and 3-month mRS outcomes as a sensitivity analysis. Blood cell traits entailed neutrophil, lymphocyte, white blood cell, and platelet counts, as well as the neutrophil to lymphocyte ratio. Genetic variants associated with blood cell counts were

ascertained from a large European GWAS by the Blood Cell Consortium X (2021) comprising over half a million individuals[31]. Genetic variants associated with neutrophil to lymphocyte ratio were derived from a UKBiobank GWAS we conducted in 340002 British participants (unpublished data; S. Methods)[32]. Causal effect estimates were expressed per 1 standard deviation increase in genetically determined blood cell traits.

## Results

### *Baseline Characteristics of INTERSTROKE cases*

A subset of 3498 stroke patients consented to genetic analysis, had peripheral blood specimen collected within one week of symptom onset, and had DNA samples that were successfully assayed for buffy coat mtDNA-CN (Supplementary Figure 1). The stroke patients analyzed in this study spanned 25 countries and 98 enrollment sites across Western Europe (26.6%), Eastern / Central Europe (11.8%), South America (28.1%), Africa (13.3%), South East Asia (6.8%), the Western Asia (6.8%), and North America and Australia (6.5%) (Table 1). The average age of stroke patients was 64.6 years (SD=14.4 years) and 1482 (42.4%) individuals were female. The sample comprised 677 (19.4%), 1259 (36.0%), and 1562 (44.6%) individuals from lower-middle / low income, upper-middle income, and high-income countries, respectively. Primary stroke types consisted of 592 (16.9%) hemorrhagic, 2889 (82.6%) ischemic, and 17 (0.5%) undefined cases. Among the 2889 patients with ischemic stroke, 54 (1.9%) had hemorrhagic transformation of their infarct. At baseline, 2010 (57.5%) participants were functionally dependent on others to perform basic activities of daily living (mRS 3-5). The level of consciousness was reduced (drowsy or unconscious) in 1129 (29.9%) patients.

**Table 1**. Demographic characteristics, comorbidities, and stroke characteristics for 3498 INTERSTROKE cases included in this study.

| **Demographic Characteristics (N=3498)** | |
|---|---|
| Age, years (SD) | 64.6 (14.4) |
| Sex, N (%) | - |
| Female | 1482 (42.4) |
| Male | 2016 (57.6) |
| Region, N (%) | - |
| Western Europe | 931 (26.6) |
| Eastern / Central Europe | 413 (11.8) |
| South America | 984 (28.1) |
| Africa | 466 (13.3) |
| South East Asia | 239 (6.8) |
| Western Asia | 238 (6.8) |
| North America / Australia | 227 (6.5) |
| Country income category, N (%) | - |
| Lower-middle or low income | 677 (19.4) |
| Upper-middle income | 1259 (36.0) |
| High income | 1562 (44.6) |
| Ethnicity, N (%) | - |
| European | 1562 (44.7) |
| Latin American | 958 (27.4) |
| African | 395 (11.3) |
| South East Asian | 259 (7.4) |
| South Asian | 108 (3.1) |
| Arab | 105 (3.0) |
| Persian | 103 (2.9) |
| Other | 8 (0.2) |
| Education, N (%) | - |
| None | 246 (7.0) |
| Primary school | 870 (24.9) |
| High school or trade school | 1545 (44.1) |
| College or university | 527 (15.1) |
| Unknown | 310 (8.9) |

| | |
|---|---|
| **Comorbidity Burden and Risk Factors** | |
| Charleson Comorbidity Index, N (%) | - |
| None | 832 (23.8) |
| One or more comorbidities | 2665 (76.2) |
| Unknown | 1 (< 0.1) |
| Risk Factors, N (%) | - |
| Hypertension | 2200 (62.9) |
| Diabetes Mellitus | 683 (19.5) |
| Hypercholesterolemia | 925 (26.4) |
| Atrial Fibrillation or Flutter | 576 (16.5) |
| Current Smoker | 776 (22.2) |
| Waist-to-hip Ratio, mean (SD) | 0.95 (0.09) |
| **Baseline Stroke Characteristics** | |
| Stroke type, N (%) | - |
| Hemorrhagic Stroke | 592 (16.9) |
| Intracerebral Hemorrhage | 587 (16.9) |
| Subarachnoid Hemorrhage | 5 (0.1) |
| Ischemic Stroke | 2889 (82.6) |
| Total anterior circulation infarct | 252 (7.2) |
| Partial anterior circulation infarct | 1333 (38.1) |
| Posterior circulation infarct | 439 (12.6) |
| Lacunar infarct | 628 (17.9) |
| Other infarct | 237 (6.8) |
| Unknown | 17 (0.5) |
| Hemorrhagic Transformation, N (%) | - |
| Present | 54 (1.9*) |
| Absent | 2835 (98.2) |
| Stroke severity, N (%) | - |
| No symptoms (mRS 0) | 151 (4.3) |
| Symptomatic but no disability (mRS 1) | 573 (16.4) |
| Slight disability (mRS 2) | 760 (21.7) |
| Moderate disability (mRS 3) | 954 (27.3) |

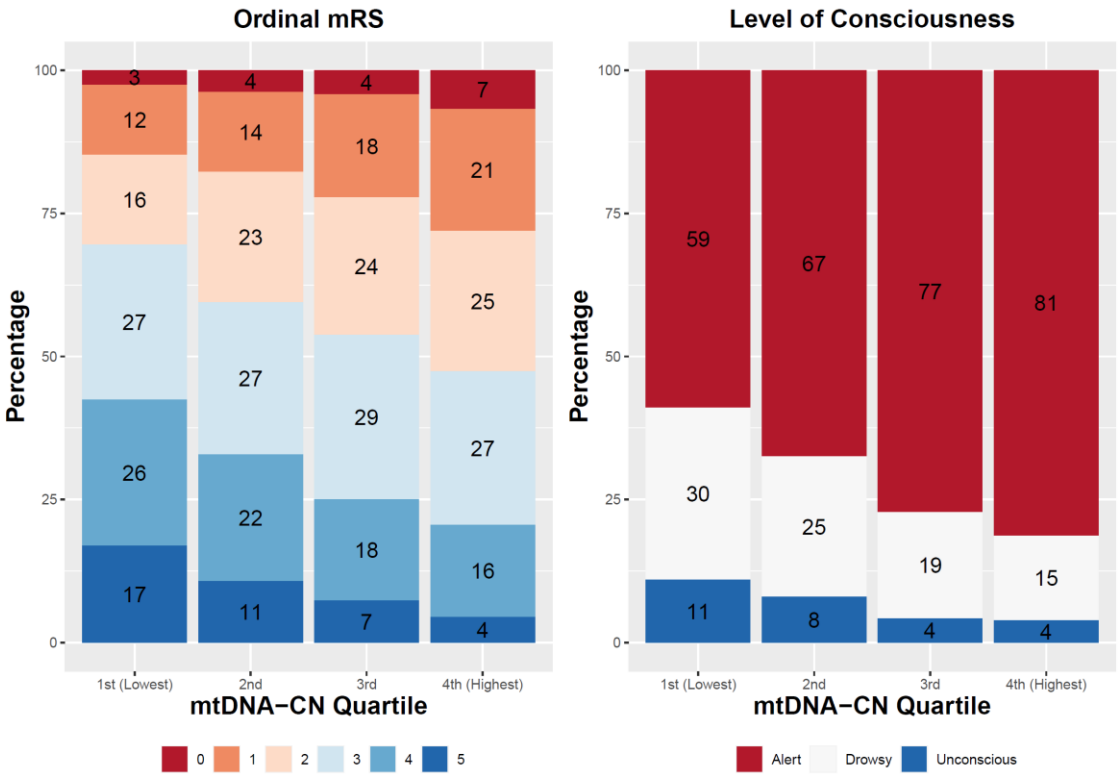
| | |
|---|---|
| Moderately severe disability (mRS 4) | 711 (20.3) |
| Severe disability (mRS 5) | 345 (9.9) |
| Unknown | 4 (0.1) |
| Level of consciousness, N (%) | - |
| Alert | 2487 (71.0) |
| Drowsy | 768 (22.0) |
| Unconscious | 237 (6.8) |
| Unknown | 6 (0.2) |

* Percentage of ischemic stroke patients, not total number of participants

***Lower mtDNA-CN is associated with greater stroke severity at baseline***

At baseline, a 1-SD lower mtDNA-CN was significantly associated with increased odds of having a more severe stroke (ordinal mRS; OR=1.27; 95% CI, 1.19-1.36; $P=4.7x10^{-12}$) and reduced consciousness (OR=1.34; 95% CI, 1.21-1.48; $P=1.8x10^{-8}$) (S. Figure 2). Among ischemic stroke patients, the association with hemorrhagic transformation non-significant (OR=1.33; 95% CI, 0.92-1.93; P=0.13). Stratifying stroke patients by mtDNA-CN quartile, there was a stepwise increase in the proportion of individuals with higher stroke severity as mtDNA-CN decreased (S. Table 1; Figure 1A). Stroke patients in the lowest mtDNA-CN quartile were at greatest risk of having a more severe stroke (OR=2.00; 95% CI,1.65-2.44; $P=2.9x10^{-12}$) and reduced consciousness (OR=2.42; 95% CI, 1.84-3.17; $P=1.6x10^{-10}$) compared to those in the highest mtDNA-CN quartile (S. Table 1; Figure 1B). These associations were step-wise and graded, and there was no significant evidence suggesting that the proportional odds assumption had been violated in any ordinal analysis (Brant $P > 0.05$; S. Table 1). Time from symptom onset to blood draw was not significantly associated with mtDNA-CN levels (P=0.11).

**A**


Ordinal mRS
Level of Consciousness
Percentage
100
75
50
25
0
3
12
16
27
26
17
4
14
23
27
22
11
4
18
24
29
18
7
7
21
25
27
16
4
59
30
11
67
25
8
77
19
4
81
15
4
1st (Lowest)
2nd
3rd
4th (Highest)
mtDNA−CN Quartile
0
1
2
3
4
5
Alert
Drowsy
Unconscious

**B**

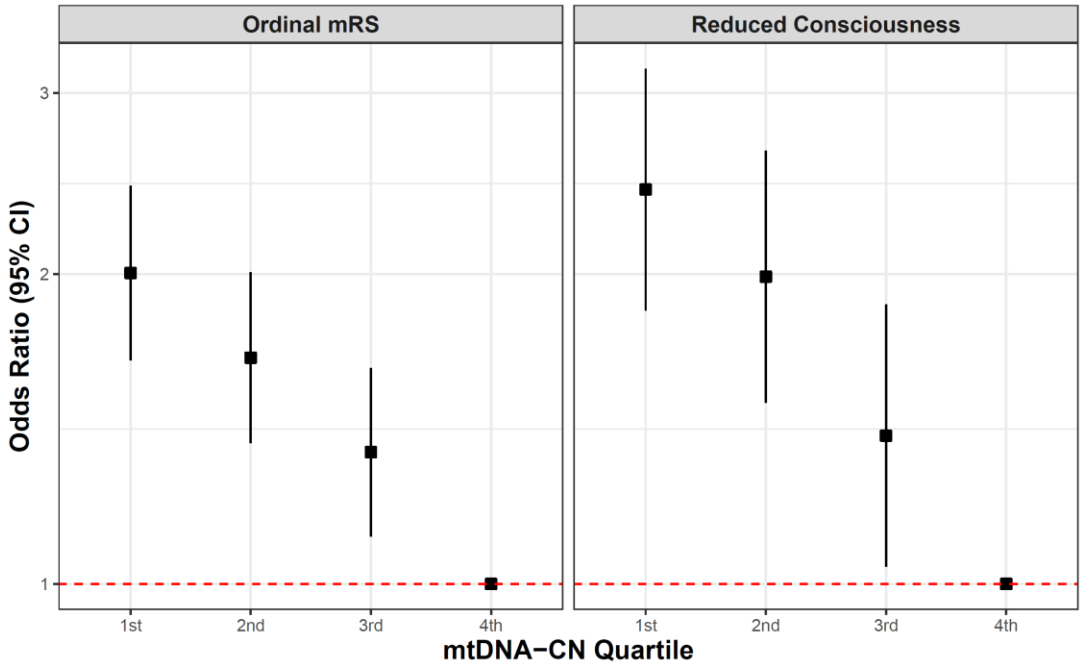

Ordinal mRS
Reduced Consciousness
Odds Ratio (95% CI)
3
2
1
1st
2nd
3rd
4th
mtDNA−CN Quartile

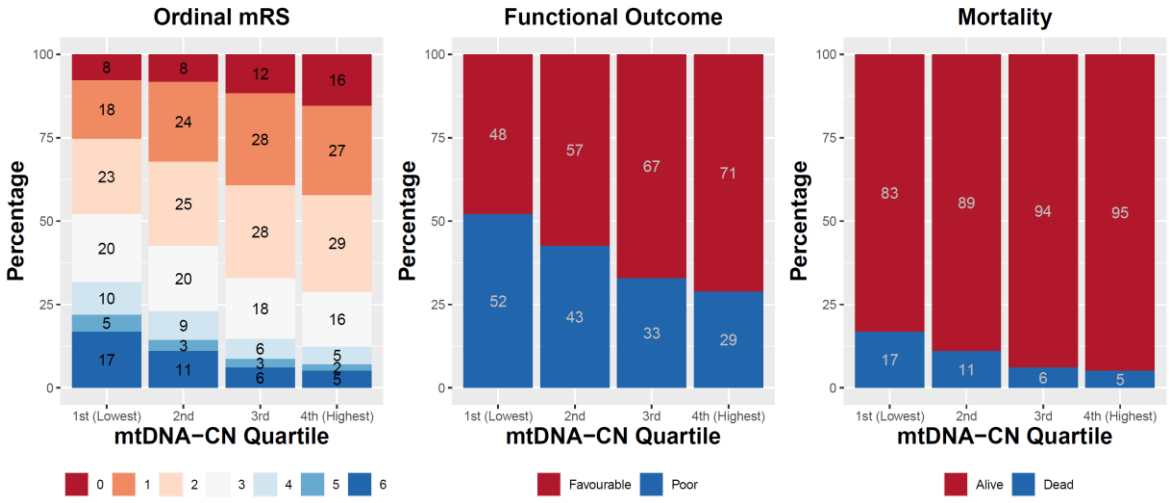
**Figure 1.** mtDNA-CN is associated with stroke severity at baseline. (A) Stacked bar plots illustrate the proportion of each (i) ordinal mRS and (ii) consciousness level category per mtDNA-CN quartile. (B) Forest plots illustrate the association between mtDNA-CN quartile and risk of having (i) more severe strokes as indicated by ordinal mRS and (ii) reduced consciousness. The highest ($4^{th}$) mtDNA-CN quartile was used as the reference group.

***Lower mtDNA-CN is associated with poor stroke prognosis at 1-month***

Of the 3498 stroke patients, mRS was recorded at follow-up for 3470 (99.2%) individuals. At 1-month follow-up, 1354 (39.0%) patients had poor functional outcome (mRS 3-6) including 337 (9.7%) patients who died. Adjusting for baseline stroke severity in addition to previous covariates, a 1-SD lower mtDNA-CN was significantly associated with higher 1-month mRS (OR=1.16; 95% CI, 1.08-1.24; $P=4.4x10^{-5}$), poor functional outcome (OR=1.21; 95% CI, 1.08-1.34; $P=6.9x10^{-4}$), and mortality (OR=1.35; 95% CI, 1.14-1.59; $P=3.9x10^{-4}$) (S. Figure 3; S. Table 2). The magnitude of effect for mtDNA-CN on mortality risk was comparable to age, an established predictor of stroke outcomes (S. Figure 3). Conversely, the effect of mtDNA-CN on post-stroke disability (mRS category and poor functional outcome status) was weaker than age (S. Figure 4). There was no significant evidence suggesting that the proportional odds assumption had been violated in any ordinal analysis (Brant $P > 0.05$; S. Table 2). Stratification by mtDNA-CN quartile revealed a consistent relationship between lower mtDNA-CN quartile and higher risk of adverse stroke outcomes (Figure 2). Stroke patients in the lowest quartile had greater odds of being classified in a higher mRS stratum (OR=1.40; 95% CI, 1.15-1.71; P=0.001),

having poor functional outcome (OR=1.51; 95% CI, 1.11-2.04; P=0.01), and mortality (OR=2.09; 95% CI, 1.34-3.25; P=0.001) compared to stroke patients in the highest quartile (S. Table 3; Figure 2).

**A**




**B**

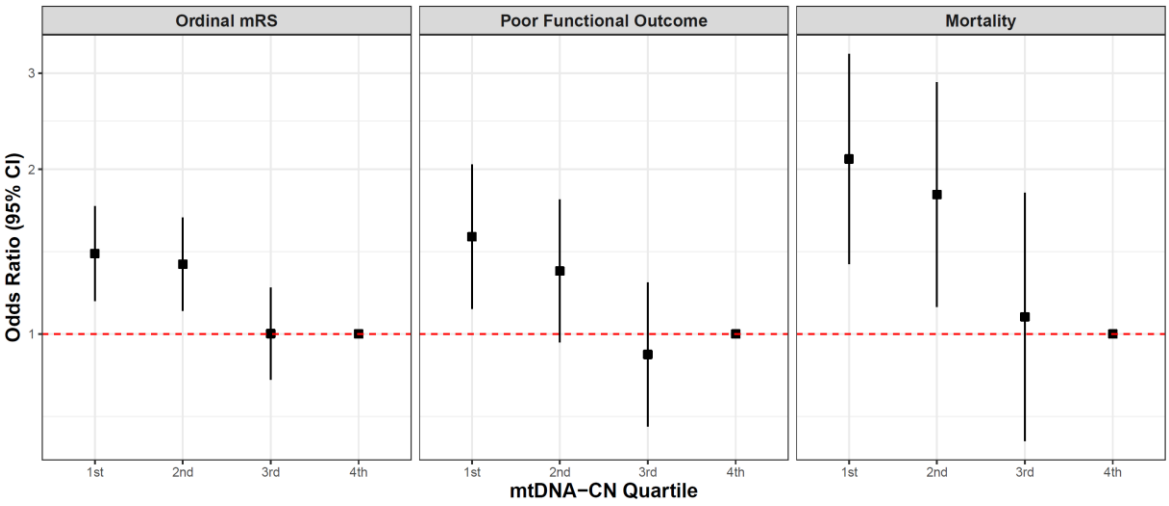


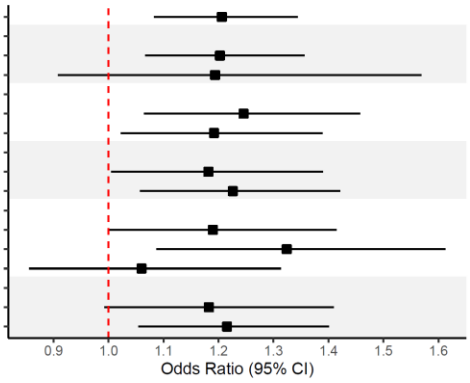

**Figure 2.** mtDNA-CN is associated with 1-month prognosis after stroke. (A) Stacked bar plots illustrate the proportion of individuals belonging to (i) ordinal mRS, (ii) functional

outcome status, and (iii) mortality categories per mtDNA-CN quartile. (B) Forest plots convey the association between mtDNA-CN quartile and post-stroke outcomes with the fourth quartile as the reference for comparison.

To further assess the robustness of mtDNA-CN-outcome associations, we performed subgroup analyses stratifying by primary stroke type, sex, age, country income level, and education level. Directionally consistent associations were observed across all subgroups for both poor functional outcome and mortality statuses with no significant heterogeneity between subgroups detected (Cochran Q Heterogeneity P > 0.10; Figure 3; S. Table 4).

**A) Poor Functional Outcome Status**

| Subgroup | N(outcome) | N(non−outcome) | OR (95% CI) | Het P |
|---|---|---|---|---|
| All | 1354 | 2116 | 1.21 (1.08−1.34) | |
| ***Primary Stroke Type*** | | | | |
| Ischemic Stroke | 916 | 1701 | 1.2 (1.07−1.36) | |
| Hemorrhagic Stroke | 303 | 245 | 1.19 (0.91−1.57) | 0.96 |
| ***Sex*** | | | | |
| Female | 638 | 828 | 1.25 (1.06−1.46) | |
| Male | 712 | 1288 | 1.19 (1.02−1.39) | 0.69 |
| ***Age*** | | | | |
| <= 65 years | 549 | 1183 | 1.18 (1−1.39) | |
| > 65 years | 801 | 933 | 1.23 (1.06−1.42) | 0.74 |
| ***Country Income Level*** | | | | |
| Lower−middle/Low | 326 | 339 | 1.19 (1−1.41) | |
| Upper−middle | 633 | 615 | 1.32 (1.09−1.61) | |
| High | 391 | 1162 | 1.06 (0.86−1.31) | 0.33 |
| ***Education Level*** | | | | |
| None or Primary | 531 | 572 | 1.18 (0.99−1.41) | |
| High School or more | 692 | 1374 | 1.22 (1.05−1.4) | 0.81 |




**B) Mortality Status**

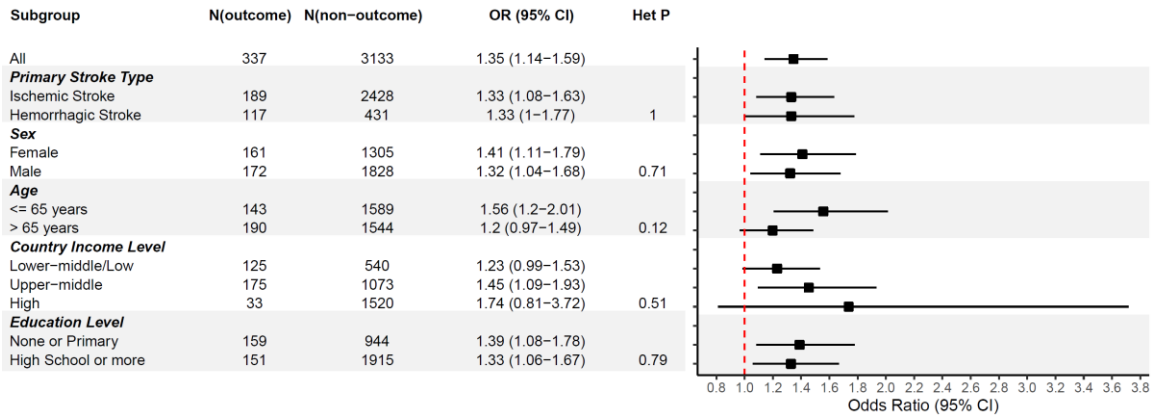



**Figure 3.** Subgroup analyses for mtDNA-CN associations with 1-month post-stroke outcomes including (A) poor functional outcome (mRS 3-6) and (B) mortality status. Except for the subgroup variable used to stratify, regression models were adjusted for age, sex, region, education level, country income level, household income level, primary stroke type and OCSP classification, Charleson comorbidity index, cardiovascular risk factors, pre-stroke disability, and baseline mRS.

Lastly, we assessed whether incorporation of mtDNA-CN improved prediction of post-stroke outcomes beyond known prognosticators, risk factors, and demographic characteristics. Addition of mtDNA-CN led to significant improvements in reclassification of functional outcome status (Net Reclassification index $(NRI)_{overall}$=0.16; 95% CI, 0.08-0.23; P=3.6x$10^{-5}$) and mortality status ($NRI_{overall}$=0.31; 95% CI, 0.19-0.43; P=1.7x$10^{-7}$). For both outcomes, NRI improvement was attributable to better reclassification of events ($NRI_{Poor\ Outcome}$=0.20; 95 % CI, 0.15-0.26; P=4.3x$10^{-12}$; $NRI_{Death}$=0.33; 95% CI, 0.22-0.44; P=3.4x$10^{-9}$) as opposed to non-events ($NRI_{Favourable\ Outcome}$=-0.05; 95% CI, -0.09 to -0.001; P=0.045; $NRI_{Alive}$=-0.02; 95% CI, -0.06 to 0.02; P=0.30) (S. Table 5).

***Low mtDNA-CN is a putative causal risk factor for 3-month stroke outcomes***

Using the UKBiobank and GISCOME studies (independent of INTERSTROKE), we found that genetically low mtDNA-CN was significantly associated with worse 3-month outcomes after stroke quantified by the ordinal mRS (OR=2.35 per SD decrease in genetically predicted mtDNA-CN; 95% CI, 1.13-4.90; P=0.02) and poor functional outcome (OR=2.68; 95% CI, 1.05-6.86; P=0.04) (Figure 4; S. Table 8). For all analyses, there was no significant evidence of directional pleiotropy (MR-Egger intercept P > 0.05), nor global heterogeneity (Cochran Q and MR-PRESSO global test P > 0.05). Results were also directionally consistent when using other MR methods (weighted median and MR-Egger) (S. Table 8). As buffy coat mtDNA-CN is known to be correlated with immune cell counts, we also performed MR analyses for blood cell traits. Despite sufficient instrument strength for neutrophil (F=100), platelet (F=154), lymphocyte (F=108), total white blood cell counts (F=106) and the neutrophil to lymphocyte ratio (F=61), none were significantly associated with 3-month outcomes (Figure 4; S. Table 9).

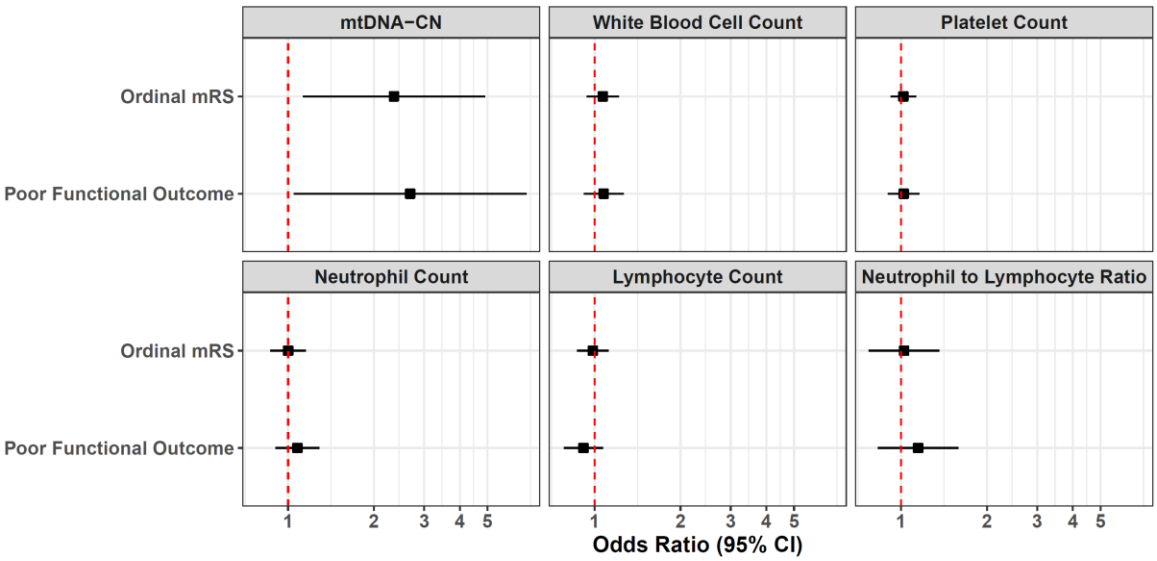


**Figure 4.** Genetic predisposition to low mtDNA-CN, but not blood cell counts, is associated with higher risk of 3-month outcomes after stroke. Effect estimates for mtDNA-CN are expressed per 1 SD decrease in genetically predicted mtDNA-CN, whereas those for blood cell traits were expressed per 1 SD increase in genetically predicted blood cell counts (or neutrophil to lymphocyte ratio). Causal effect estimates obtained by the inverse variance weighted method are displayed as there was no significant heterogeneity or directional pleiotropy detected for any analysis (S. Tables 6 & 7).

## Discussion

Our study represents the first international multicenter exploration of buffy coat mtDNA-CN as a potential prognosticator of post-stroke outcomes. First, lower buffy coat mtDNA-CN measured within one week post symptom onset correlated with functional and clinically relevant stroke severity indicators such as higher mRS and reduced consciousness. Second, lower buffy coat mtDNA-CN was associated with greater risk of poor functional outcome and death at 1-month follow-up, which were consistent across primary stroke type, sex, age, country income level, and education level strata, as well as, independent of baseline stroke severity. Third, in addition to being a strong predictor of mortality with a magnitude of effect comparable to, if not stronger than, age, the inclusion of buffy coat mtDNA-CN improved the prediction of functional outcome and death. Fourth, MR analysis provided support for low buffy coat mtDNA-CN as a causal mediator of 3-month mRS and poor functional outcome status. Altogether, our findings confirm the hypothesis that low buffy coat mtDNA-CN is a biomarker and mediator of worse stroke prognosis.

The main clinical implication of our study is that buffy coat mtDNA-CN may represent a useful prognostic marker of post-stroke outcomes. First, buffy coat mtDNA-CN is a blood biomarker of post-stroke outcomes that does not suffer from inter-rater variability and is not influenced by a patient's communication deficit. Second, the mtDNA-CN-outcome associations are consistent across stroke type, sex, age, country income level, education level, and baseline severity, which positions mtDNA-CN to have widespread utility across stroke patients globally. To our knowledge, we provide the first evidence suggesting that low mtDNA-CN may have consistent effects in both ischemic and hemorrhagic stroke patients. This is particularly relevant for health systems in low-income settings, which bear a disproportionate global burden of hemorrhagic stroke[1–3], though further analyses in larger samples of hemorrhagic stroke patients are warranted to confirm. Third, low mtDNA-CN represents a strong risk marker with effects comparable to established prognosticators including older age. Moreover, the observed effect for mtDNA-CN on mortality is also comparable to that of carrying an *APOE* Ɛ2 allele, which confers a 1.5-fold increased risk of 3-month mortality in intracerebral hemorrhage patients and is present in approximately 15% of the population[34]. For comparison, we found that stroke patients in the bottom 15% of mtDNA-CN levels had a 1.6-fold increased risk of 1-month mortality (OR=1.57; 95% CI, 1.13-2.17; P=0.007) relative to the remaining 85% participants with higher mtDNA-CN levels. Fourth, mtDNA-CN is an easily accessible biomarker as (i) it can be measured from peripheral blood after stroke, (ii) the assay necessitates only basic molecular laboratory techniques (qPCR), and (iii) the cost per sample is low (< $5 USD). Logistic and operational convenience combined with evidence

for robust, objective, and strong prognostic utility raises the promising prospect of implementing mtDNA-CN clinically; however, replication of such findings in a prospective analysis in addition to formal economic analyses in various settings is warranted.

Findings from MR analyses suggest that proper mtDNA regulation may be imperative for stroke protection and recovery, which aligns with animal model experiments demonstrating an important role for mtDNA-CN regulators in mediating protection against ischemia reperfusion injury. For example, reoxygenation of rodents with acute kidney injury induces the formation of excessive mitochondrial reactive oxygen species, accompanied by a sharp decline in mtDNA-CN levels[24]. In addition, genetic upregulation of the mtDNA replication initiation factor, TFAM, is sufficient to rescue this acute drop in mtDNA-CN levels thereby attenuating ischemia reperfusion injury. In the context of stroke models, mice with transient middle cerebral artery occlusion exhibit excessive cleavage of OPA1, another important mtDNA regulator, and treatment with either a cleavage-resistant form of OPA1 or mild overexpression of OPA1 markedly reduces infarct volume and neuronal apoptosis[23,35]. In conjunction with prior mechanistic studies, our epidemiological and genetic findings contribute to the mounting evidence that maintaining adequate mtDNA-CN may mediate cellular resilience to ischemic insults. Furthermore, consistent epidemiological associations in hemorrhagic stroke patients suggest that mtDNA-CN may protect against stroke injury through general mechanisms pertinent to both etiologies (e.g. blood brain barrier disruption, neuroprotection, inflammation, etc.)[36]. Future MR analyses

and experiments are necessary to elucidate the effects of mtDNA-CN perturbation on post-stroke outcomes in the hemorrhagic context specifically.

Our study had several limitations. First, as INTERSTROKE was a large international case-control study, measures of baseline stroke severity (NIHSS) and outcome (3-month mRS) that are common in smaller stroke research studies were substituted with baseline mRS and 1-month mRS for feasibility, respectively, as was done in Langhorne *et al.* (2018). The interchangeability of such measures has been validated in previous independent studies showing high correlation between baseline NIHSS and mRS (r=0.69) and between 1-month and 3-month mRS (r=0.87; weighted kappa agreement = 0.86)[37,38]. Second, complete blood cell counts were not measured in INTERSTROKE participants; thus, we cannot directly evaluate to what extent blood cell counts influence observational associations with post-stroke outcomes. However, our genetic analyses suggest that mtDNA-CN may have a direct role in stroke prognosis independent of changes in blood cell counts since (i) mtDNA-CN GWAS effects had already been adjusted for major cell count determinants of mtDNA-CN levels (neutrophil, white blood cell, and platelet counts) and (ii) no significant association was observed for genetically determined immune cell counts *per se*. Nonetheless, the genetic determinants of post-stroke immune cell changes may differ from those influencing variation in cell counts within the general population as suggested by results from Torres-Aguila *et al.* (2019)[39]. Third, although associations were corrected for a crude surrogate of infarct volume (OCSP classification), direct measurements of infarct and hematoma volumes were not available. Fourth, survivorship bias may have led to conservative effect estimates as INTERSTROKE cases

included patients surviving to hospital admission, and consequently, patients with severe, early fatal strokes were not represented. Finally, MR analyses were limited by the following considerations: (i) causal effect estimates were imprecise and were accompanied by large confidence intervals though these were consistent in direction-of-effect with epidemiological associations, (ii) although sensitivity analyses did not show significant evidence of heterogeneity, directional pleiotropy, or outlying effects, it is impossible to completely exclude bias due to potential pleiotropy, (iii) mortality and hemorrhagic stroke outcomes could not be evaluated directly for lack of GWAS summary statistics, and (iv) analyses were solely based on European participants.

## Conclusions

Low buffy coat mtDNA-CN measured within one week of symptom onset represents an accessible and robust biomarker of both stroke severity and prognosis. MR findings suggest that low mtDNA-CN may mediate post-stroke outcomes. Additional investigations are warranted to replicate such findings in additional populations, to establish the temporal profile of post-stroke mtDNA-CN changes in more detail, and to assess whether compounds that maintain mtDNA-CN levels after cerebral insult hold promise as a novel therapeutic strategy.

## Acknowledgements

We would like to acknowledge the investigators and participants of the INTERSTROKE, UKBiobank, and GISCOME studies for their important contributions to this research.

## Competing Interests

We declare no competing interests.

## Author Contributions

Mr. Chong and Dr. Pare had full access to all of the data and take responsibility for the integrity of the data and the accuracy of the data analysis.

*Study concept and design:* Chong, Paré.

*Acquisition, analysis, or interpretation of data:* Chong, Langhorne, Narula, Morton, Judge, Akhabir, Cawte, Pathan, Lali, Mohammadi-Shemirani, Shoamanesh, O'Donnell, Yusuf, Paré.

*Drafting of the manuscript:* Chong.

*Critical revision of the manuscript for important intellectual content:* All authors.

*Statistical analysis:* Chong.

*Obtained funding:* O'Donnell, Yusuf, Paré.

*Administrative, technical, or material support:* Judge, O'Donnell, Langhorne

*Supervision:* Paré.

## Funding

*Study*

Canadian Institutes of Health Research (399497), Canadian Stroke Network, Heart and Stroke Foundation Canada (G-18–0022359)

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication

*Co-Authors*

Canadian Institutes of Health Research Frederick Banting and Charles Best Canada Graduate Scholarships Doctoral Award (MC, PM)

Canadian Institutes of Health Research Post-Doctoral Fellowship Award (RM)

Wellcome Trust Health Research Board Irish Clinical Academic Training (ICAT) Programme Grant Number: 203930/B/16/Z (CJ)

European Research Council COSIP Grant Number: 640580 (MO)

Chest, Heart and Stroke Scotland (PL)

CISCO Professorship in Integrated Health Systems (GP)

Canada Research Chair in Genetic and Molecular Epidemiology (GP)

**Conflict of Interest:** None

**Research Ethics Statement:** Research was approved by the Hamilton Integrated Research Ethics Board under project # 06-331. INTERSTROKE analyses were reported following STREGA guidelines, and MR analyses were reported according to STROBE-MR guidelines (S. Tables 10 and 11).

**Data Availability:** The principal and corresponding authors, Michael Chong and Guillaume Paré, take full responsibility for the data, the analyses and interpretation, and the conduct of the research; all authors have full access to all of the data; and all authors have the right to publish any and all data separate and apart from any sponsor. The primary datasets in this study include INTERSTROKE, UKBiobank, and GISCOME. INTERSTROKE data may be available upon approval for request of collaboration with study PI, Martin O'Donnell. UKBiobank individual-level data can be acquired upon application (https://bbams.ndph.ox.ac.uk/ams/). UKBiobank mtDNA-CN GWAS

summary statistics will be posted on Github (https://github.com/GMELab). GISCOME summary statistics are freely available to download (https://cd.hugeamp.org/downloads.html).

**References**


1. Langhorne P, O'Donnell MJ, Chin SL, et al. Practice patterns and outcomes after stroke across countries at different economic levels (INTERSTROKE): an international observational study. *Lancet*. 2018;391(10134):2019-2027. doi:10.1016/S0140-6736(18)30802-X
2. Yusuf S, Rangarajan S, Teo K, et al. Cardiovascular Risk and Events in 17 Low-, Middle-, and High-Income Countries. *N Engl J Med*. 2014;371(9):818-827. doi:10.1056/NEJMoa1311890
3. Dagenais GR, Leong DP, Rangarajan S, et al. Variations in common diseases, hospital admissions, and deaths in middle-aged adults in 21 countries from five continents (PURE): a prospective cohort study. *Lancet*. 2020;395(10226):785-794. doi:10.1016/S0140-6736(19)32007-0
4. Montaner J, Ramiro L, Simats A, et al. Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke. *Nat Rev Neurol*. 2020;16(5):247-264. doi:10.1038/s41582-020-0350-6
5. Liu F, Lu J, Manaenko A, Tang J, Hu Q. Mitochondria in ischemic stroke: New insight and implications. *Aging Dis*. 2018;9(5):924-937. doi:10.14336/AD.2017.1126
6. Li Q, Gao S. Mitochondrial Dysfunction in Ischemic Stroke. In: *Translational*

*Research in Stroke*. ; 2017:201-221. doi:10.1007/978-981-10-5804-2_10

7. Paramasivam A, Venkatapathi C, Sandeep G, et al. Homozygous R627W mutations in POLG cause mitochondrial DNA depletion leading to encephalopathy, seizures and stroke-like episodes. *Mitochondrion*. 2019;48(June 2018):78-83. doi:10.1016/j.mito.2019.08.003
8. Bonora E, Chakrabarty S, Kellaris G, et al. Biallelic variants in LIG3 cause a novel mitochondrial neurogastrointestinal encephalomyopathy. *Brain*. Published online 2021:1-17. doi:10.1093/brain/awab056
9. Malik AN, Czajka A. Is mitochondrial DNA content a potential biomarker of mitochondrial dysfunction? *Mitochondrion*. 2013;13(5):481-492. doi:10.1016/j.mito.2012.10.011
10. Koller A, Fazzini F, Lamina C, et al. Mitochondrial DNA copy number is associated with all-cause mortality and cardiovascular events in patients with peripheral arterial disease. *J Intern Med*. 2020;287(5):569-579. doi:10.1111/joim.13027
11. Zhang Y, Guallar E, Ashar FN, et al. Association between mitochondrial DNA copy number and sudden cardiac death : findings from the Atherosclerosis Risk in Communities study ( ARIC ). Published online 2017:3443-3448. doi:10.1093/eurheartj/ehx354
12. Fazzini F, Lamina C, Fendt L, et al. Mitochondrial DNA copy number is associated with mortality and infections in a large cohort of patients with chronic kidney disease. 2019;8:480-488. doi:10.1016/j.kint.2019.04.021
13. Song L, Liu T, Song Y, et al. mtDNA Copy Number Contributes to All-Cause

Mortality of Lacunar Infarct in a Chinese Prospective Stroke Population. *J Cardiovasc Transl Res*. 2020;13(5):783-789. doi:10.1007/s12265-019-09943-9

14. O ’donnell MJ, Chin SL, Rangarajan S, et al. Global and regional eff ects of potentially modifi able risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. *Lancet*. 2016;388(20):761-775. doi:10.1016/S0140-6736(16)30506-2
15. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
16. Soderholm M, Pedersen A, Lorentzen E, et al. Genome-wide association meta-analysis of functional outcome after ischemic stroke. *Neurology*. 2019;0:1271-1283. doi:10.1212/WNL.0000000000007138
17. O’Donnell MJ, Xavier D, Liu L, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet*. 2010;376(9735):112-123. doi:10.1016/S0140-6736(10)60834-3
18. Uyttenboogaart M, Stewart RE, Vroomen PCAJ. Optimizing Cutoff Scores for the Barthel Index and the Modified Rankin Scale for Defining Outcome in Acute Stroke Trials. Published online 2005:1984-1987. doi:10.1161/01.STR.0000177872.87960.61
19. Fazzini F, Schöpf B, Blatzer M, et al. Plasmid-normalized quantification of relative mitochondrial DNA copy number. 2018;(May):1-11. doi:10.1038/s41598-018-33684-5

20. Smith GD, Hemani G. Mendelian randomization: Geneticanchorsfor causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23(R1):89-98. doi:10.1093/hmg/ddu328

21. Hemani G, Zheng J, Elsworth B, et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:1-29. doi:10.7554/eLife.34408

22. Bandres-Ciga S, Noyce AJ, Traynor BJ. Mendelian Randomization—A Journey From Obscurity to Center Stage With a Few Potholes Along the Way. *JAMA Neurol*. 2020;77(1):7. doi:10.1001/jamaneurol.2019.3419

23. Lai Y, Lin P, Chen M, et al. Restoration of L-OPA1 alleviates acute ischemic stroke injury in rats via inhibiting neuronal apoptosis and preserving mitochondrial function. *Redox Biol*. 2020;34(March):101503. doi:10.1016/j.redox.2020.101503

24. Zhao M, Wang Y, Li L, et al. Mitochondrial ROS promote mitochondrial dysfunction and inflammation in ischemic acute kidney injury by disrupting TFAM-mediated mtDNA maintenance. *Theranostics*. 2021;11(4). doi:10.7150/thno.50905

25. Chong M, Mohammadi-Shemirani P, Perrot N, et al. GWAS and ExWAS of blood Mitochondrial DNA copy number identifies 73 loci and highlights a potential causal role in dementia. *medRxiv*. Published online 2021. doi:10.1101/2021.04.08.21255031

26. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779

27. Rath S, Sharma R, Gupta R, et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res*. 2021;49(D1):D1541-D1547. doi:10.1093/nar/gkaa1011

28. Kamat MA, Blackshaw JA, Young R, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics*. 2019;35(22):4851-4853. doi:10.1093/bioinformatics/btz469

29. Moore AZ, Ding J, Tuke MA, et al. Influence of cell distribution and diabetes status on the association between mitochondrial DNA copy number and aging phenotypes in the InCHIANTI study. *Aging Cell*. 2018;17(1):6-8. doi:10.1111/acel.12683

30. Hurtado-Roca Y, Ledesma M, Gonzalez-Lazaro M, et al. Adjusting MtDNA quantification in whole blood for peripheral blood platelet and leukocyte counts. *PLoS One*. 2016;11(10):1-14. doi:10.1371/journal.pone.0163770

31. Vuckovic D, Bao EL, Akbari P, et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell*. 2020;182(5):1214-1231.e11. doi:10.1016/j.cell.2020.08.008

32. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole genome regression for quantitative and binary traits.

33. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50(5):693-698. doi:10.1038/s41588-018-0099-7

34. Biffi A, Anderson CD, Jagiella JM, et al. APOE genotype and extent of bleeding

and outcome in lobar intracerebral haemorrhage : a genetic association study. *Lancet Neurol*. 2011;10(8):702-709. doi:10.1016/S1474-4422(11)70148-X

35. Varanita T, Soriano ME, Sandri M, et al. The Opa1-Dependent Mitochondrial Cristae Remodeling Pathway Controls Atrophic , Apoptotic , Article The Opa1-Dependent Mitochondrial Cristae Remodeling Pathway Controls Atrophic , Apoptotic , and Ischemic Tissue Damage. Published online 2015:834-844. doi:10.1016/j.cmet.2015.05.007
36. Chen W, Guo C, Feng H, Chen Y. Mitochondria: Novel Mechanisms and Therapeutic Targets for Secondary Brain Injury After Intracerebral Hemorrhage. *Front Aging Neurosci*. 2021;12(January):1-10. doi:10.3389/fnagi.2020.615451
37. Bruno A, Close B, Switzer JA, et al. Simplified modified Rankin Scale questionnaire correlates with stroke severity. *Clin Rehabil*. 2013;27(8):724-727. doi:10.1177/0269215512470674
38. Ovbiagele B, Lyden PD, Saver JL. Disability status at 1 month is a reliable proxy for final ischemic stroke outcome. *Neurology*. 2010;44364:688-692.
39. Torres-Aguila NP, Carrera C, Giese AK, et al. Genome-wide association study of white blood cell counts in patients with ischemic stroke. *Stroke*. 2019;50(12):3618-3621. doi:10.1161/STROKEAHA.119.026593

# CHAPTER 6:

# DISCUSSION

# CHAPTER 6: DISCUSSION

## 6.1 GENERAL OVERVIEW

Despite current treatments, stroke risk remains high thus emphasizing the need to discover new drug targets. By their very definition, stroke biomarkers have clinical utility for stroke risk assessment, prognostication, or treatment response, but their latent value is that a subset of biomarkers are causal mediators of disease, thereby representing putative therapeutic targets. Throughout this thesis, we used MR analysis to identify circulating markers that also represent putative causal mediators. In Chapter 3, we systematically interrogated the blood proteome and uncovered established drug targets and novel putative mediators. Chapters 4 and 5 were devoted to investigation of an emerging mitochondrial biomarker, leukocyte mtDNA-CN. Specifically, Chapter 4 provided insights into the genetic architecture of leukocyte mtDNA-CN and fulfilled the first step of MR, which is to identify a subset of genetic variants associated with exposure in question. Using these identified mtDNA-CN-qTLs, Chapter 5 elucidated a role for mtDNA-CN as a predictor and causal risk factor for post-stroke outcomes. Altogether, this thesis identifies several potential therapeutic targets for stroke using the MR framework. In the subsequent sections, we will (i) highlight the main findings from each chapter, (ii) outline the biological, clinical, and research implications of these works, (iii) discuss new relevant literature and additional considerations to contextualize findings, and (iv) describe strengths, limitations, and future areas of investigation.

## 6.2 CHAPTER SUMMARIES

### 6.2.1 STUDY 1 (CHAPTER 3) SUMMARY

A systematic MR screen of 653 circulating proteins identified putative causal roles for previously established (ABO, CD40, FXI, LP(a), MMP12) and novel (SCARA5, TNFSF12) proteins mediating ischemic stroke risk. Forecasting potential adverse side-effects, MR analysis revealed that if TNFSF12 was therapeutically targeted for ischemic stroke reduction, then this would also be accompanied by increased intracranial bleeding. Agnostic phenome-wide MR analyses identified 71 secondary associations with diseases, indicating substantial pleiotropic effects. Phenome-wide MR findings provided some reassurance of the safety of emerging therapeutic targets, FXI and LPA, for which inhibitors are currently undergoing phase III RCTs. Lastly, MR analyses suggest that SCARA5 may be a promising therapeutic target for treatment of cardioembolic stroke with no adverse side-effects on intracranial bleeding or other diseases detected.

### 6.2.3 STUDY 2 (CHAPTER 4) SUMMARY

To enable large-scale genetic investigations of mtDNA-CN, we first developed and validated a novel method, “AutoMitoC”, to infer mtDNA-CN from widely accessible genetic array data and then applied it to 395,781 participants from the UKBiobank study. Genetic analyses (genome-wide and exome-wide association studies) identified 71 loci, implicating genes involved in rare mtDNA depletion disorders, dNTP metabolism, and the mitochondrial central dogma. mtDNA depletion syndromes are rare genetic disorders characterized by severely low mtDNA-CN, and novel strategies to increase mtDNA levels may benefit these patients who currently have no treatments. Analysis of rare variants

revealed that rare protein-altering mutations in *SAMHD1* were associated with higher mtDNA-CN levels, thus representing a potential therapeutic target for mtDNA depletion. Conversely, we also found that *SAMHD1* mutation carriers had approximately two-fold greater risk of breast cancer. Finally, MR analysis identified a causal relationship between genetically low mtDNA-CN and increased risk of dementia.

### 6.2.4 STUDY 3 (CHAPTER 5) SUMMARY

We performed the first global characterization of mtDNA-CN as a prognosticator for stroke outcomes in 3,498 stroke patients from the INTERSTROKE study. Lower mtDNA-CN measured within one week of symptom onset correlated with several stroke severity indicators at baseline. Independent of baseline stroke severity, lower mtDNA-CN was also associated with worse prognosis at 1-month including higher risk of poor functional outcome (mRS 3-6 vs. 0-2) and mortality. Notably, the magnitude of the mtDNA-CN effect was comparable to that of established prognosticators, such as older age and *APOE* ε2 carrier status. Addition of mtDNA-CN to statistical models significantly improved event reclassification of poor functional outcome and mortality. MR analysis supported a causal relationship between genetically low mtDNA-CN and worse 3-month functional outcomes.

## 6.3 SIGNIFICANCE OF FINDINGS

### 6.3.1 CLINICAL IMPLICATIONS

The major clinical significance of this work is the identification of several novel therapeutic candidates to potentially ameliorate stroke risk and progression. Chapter 3 revealed circulating TNFSF12 and SCARA5 as putative causal mediators of cardioembolic

stroke. Chapter 5 uncovered a causal link between low leukocyte mtDNA-CN and worse post-stroke outcomes, implying that interventions that upregulate or maintain mtDNA-CN levels during stroke may help reduce stroke severity and improve functional recovery. For example, SAMHD1 inhibition as suggested by Chapter 4, could be a means of recovering mtDNA-CN levels. Altogether, such findings represent suggestive pre-clinical human genetics evidence supporting a causal role of these molecules for stroke risk and outcomes. Accordingly, substantial follow-up investigations are necessary to better understand the underlying causative tissues and mechanisms that mediate the observed biomarker-stroke relationships, as well as to assess practical aspects of drug development including druggability, bioavailability, and drug delivery. Lastly, we also found that blood-based mtDNA-CN is an accessible, robust, and objective marker of stroke severity and progression in a globally representative sample of acute stroke cases. While it is tempting to speculate that this marker may have widespread utility for risk stratification and disease progression tracking, additional studies are necessary to assess whether mtDNA-CN provides complementary prognostic utility to established and emerging markers, such as cerebral microbleeds and circulating NFl. Lastly, we provide the first human genetics evidence directly linking dysregulation of mtDNA-CN to worsening post-stroke outcomes.

### 6.3.2 BIOLOGICAL IMPLICATIONS

This work reinforces the prevailing notion that stroke biology is highly complex. Numerous pathways are implicated by the identified causal mediators including atherosclerosis (LP(a), CD40), thrombosis (ABO, FXI, LP(a), SCARA5), inflammation (ABO, LP(a), TNFSF12), vascular and atrial remodeling (MMP12, TNFSF12), iron

metabolism (SCARA5, TNFSF12), and mitochondrial dysfunction (mtDNA-CN). Many of these proteins participate in multiple processes relevant to stroke thus reflecting the multifunctionality of individual proteins and widespread pleiotropy.

Similarly, GWAS and ExWAS findings highlight the complexity of mtDNA-CN as a mitochondrial biomarker. While historically perceived by some as a simple surrogate measure to quantify cellular mitochondria content, genetic association results provide a more nuanced understanding of what this biomarker represents implicating multiple mitochondrial processes including mtDNA central dogma, nucleotide supply and metabolism, mitochondrial respiration, mitochondrial biogenesis, and other mitochondrial dynamics[1]. Additionally, GWAS results affirm a polygenic basis for mtDNA-CN since we identified 71 genetic loci. To put this into perspective, this represents a 40% increase in our knowledge of mtDNA-CN loci as compared to the next largest published GWAS[2]. The polygenicity of mtDNA-CN also implies perhaps that milder polygenic forms of mtDNA depletion may exist, though future studies are required to understand the aggregate effects of polygenic dysregulation of mtDNA-CN and the mechanisms responsible for variant and gene-based associations.

### 6.3.3. RESEARCH IMPLICATIONS

The frameworks, tools, and knowledge contributed by this thesis may help guide future drug target prioritization initiatives for cerebrovascular disease, cognitive decline, mitochondrial disorders, and other conditions. Chapter 3 conveyed a framework for how MR can be broadly applied to several aspects of drug target evaluation, ranging from systematic identification of candidate drug targets, safety adjudication and the elucidation

of unexpected side-effects, and understanding mediating risk factors. Chapter 4 described the primary methodological contribution of this thesis, namely, the development of AutoMitoC, a pipeline to infer mtDNA-CN from SNP microarray intensities for large multi-ethnic biobank studies. Our hope is that AutoMitoC serves the broader mitochondria research community by enabling international collaborations to further interrogate the genetic determinants of mtDNA-CN. Lastly, genome-wide summary statistics for the UKBiobank mtDNA-CN GWAS will be made publicly available. This should facilitate additional MR analyses to identify other complications of mtDNA-CN dysregulation akin to what was performed for post-stroke outcomes in Chapter 5, and to find novel interventions to modulate mtDNA-CN levels.

## 6.4 DISCUSSION OF PUTATIVE STROKE TARGETS

### 6.4.1 SCAVENGER CLASS A RECEPTOR MEMBER A5 (SCARA5)

SCARA5 is an endocytic scavenger receptor that is widely expressed in the epithelium of various tissues[3,4]. As a class A endocytic receptor, SCARA5 mediates the clearance of a wide repertoire of compounds from the stroma, and thus has important roles in various biological processes including tumour suppression, innate immunity, iron homeostasis, and hemostasis[3–10].

#### 6.4.1.1 NEW STUDIES IMPLICATE SCARA5 AS A THROMBOSIS REGULATOR

In Study 1 (Chapter 3), the link between circulating SCARA5 and cardioembolic stroke risk was speculated to be due to its function as a transporter of L-ferritin since circulating iron is a causal risk factor for cardioembolic stroke. While this remains plausible, new evidence supports a role for SCARA5 in coagulation. Multiple GWAS

analyses show that an intronic SCARA5 cis-pQTL (rs2726927) is associated with venous thromboembolism susceptibility, activated partial thromboplastin time (a measure of clotting time), coagulation factor VIII (FVIII) levels, and von Willebrand Factor (vWF) levels[9,11,12]. Circulating vWF has a prominent role in hemostasis as a stabilizer of FVIII and an anchor for platelet adhesion to damaged parts of the endothelium. Recent experimental evidence also corroborates a direct interaction between SCARA5 and vWF with SCARA5 acting as an endocytic receptor for vWF. Specifically, *in vivo* and *in vitro* rodent experiments by Swystun *et al.* (2019) found that SCARA5 expressed on the surface of splenic littoral cells facilitated the clearance of circulating vWF and FVIII (through vWF-FVIII complexes)[10]. Altogether, upregulation of SCARA5's endocytic receptor activity may represent a novel therapeutic strategy for mitigating cardioembolic stroke.

#### 6.4.1.2 ADDITIONAL CONSIDERATIONS FOR THE INTERPETATION OF SCARA5 MR RESULTS

While our UKBiobank-based MR analyses did not uncover side-effects of genetic upregulation of SCARA5 for intracranial bleeding nor other diseases, further investigations are required because (i) the MR analyses for SCARA5 were less well-powered in comparison to other circulating mediators like LP(a) and thus it is possible that we lacked adequate power to detect association with bleeding phenotypes, (ii) if stroke protection conferred by SCARA5 upregulation is indeed mediated through enhanced clearance of vWF, then bleeding risk may be concerning as vWF deficiency causes bleeding diathesis, and (iii) if stroke protection is partially mediated by removal of iron from circulation, then an increased risk of anemia might be expected. As a counterargument to the second point,

recombinant vWF is available as an antidote for bleeding diathesis. Finally, it is important to consider that the main MR finding implicated circulating SCARA5 levels, but the former research literature has almost exclusively interrogated the endocytic activity of the membrane-bound form[13].

### 6.4.2 TUMOR NECROSIS FACTOR LIGAND SUPERFAMILY MEMBER 12 (TNFSF12)

In Study 1 (Chapter 3), antagonistic effects were observed for the associations between genetically determined TNFSF12 levels on ischemic and hemorrhagic stroke subtypes. Specifically, higher TNFSF12 was associated with decreased risk of cardioembolic stroke but increased risk of intracranial bleeding. Circulating TNFSF12 is a multifunctional cytokine that acts primarily by binding to its membrane-bound receptor, Fibroblast Growth Factor-Inducible 14 (Fn14)[14]. Activation of the TNFSF12/Fn14 axis leads to the initiation of deleterious cellular signalling pathways including inflammation (IL6), cell proliferation and infiltration (PI3K-AKT and MKK-ERK1/2), and extracellular matrix remodeling (MMPs)[14].

#### 6.4.2.1 EMERGING EVIDENCE FOR TNFSF12 AS A PROGNOSTICATOR OF POST-STROKE OUTCOME

Independent of hypertension, a recent series of investigations by Silva-Candal *et al.* (2020; 2021) also suggest prognostic utility for circulating TNFSF12 levels as a biomarker of acute and post-stroke outcomes[15,16]. These studies observed that higher circulating TNFSF12 at hospital admission is associated with increased risks of (i) hemorrhagic transformation among ischemic stroke patients, (ii) early hematoma expansion in

intracerebral hemorrhage patients, and (iii) poor functional outcome 3-months after stroke in both ischemic and hemorrhagic stroke patients[15,16]. Accordingly, these new associations imply that the relationship between TNFSF12 and intracranial bleeding may be mediated by additional mechanisms beyond hypertension alone which warrants further investigation.

#### 6.4.2.2 ADDITIONAL CONSIDERATIONS FOR THE INTERPETATION OF TNFSF12 MR RESULTS

Our phenome-wide MR analysis hinted that TNFSF12 may mediate stroke through atrial fibrillation and hypertension. Indeed, a recent proteome-wide MR analysis for atrial fibrillation replicated the causal protective effects of circulating TNFSF12 using an independent dataset for atrial fibrillation[17]. In terms of validation efforts for the relationship between TNFSF12 and hypertension, this finding has yet to be replicated in an independent dataset; however, it was reproduced in the same UKBiobank dataset by another research group using different MR parameters (https://www.epigraphdb.org/pqtl/TNFSF12;TNFSF12-TNFSF13).

TNFSF12 upregulation has not been explored therapeutically, but RCTs have assessed the safety and clinical efficacy of TNFSF12/Fn14 axis blockade with monoclonal antibodies (BIIB023, RO-5458640) in the contexts of cancer, lupus nephritis, and rheumatoid arthritis[18–20]. While these studies were prematurely terminated due to a lack of clinical efficacy, there is renewed enthusiasm for TNFSF12/Fn14 inhibition for treatment of cardiovascular disease [14]. Compelling evidence comes from Méndez-Barbero *et al.* (2019) who observed that TNFSF12 is highly expressed in human coronary arteries with in-stent restenosis, and that pharmacologic inhibition of the TNFSF12/Fn14 axis

ameliorates post-angioplasty restenosis in mice[21]. Although it is enticing to speculate about drug repurposing opportunities, importantly, our MR findings imply a potential adverse effect of TNFSF12 blockade on atrial fibrillation risk. Moreover, given that the target population would consist of cardiovascular disease patients, the concomitant cost of increasing risk of a major cardiovascular risk factor, atrial fibrillation, should be carefully considered.

### 6.4.3 MITOCHONDRIAL DNA COPY NUMBER (MTDNA-CN)

#### 6.4.3.1 MTDNA-CN RECOVERY AS A POTENTIAL THERAPEUTIC AXIS FOR CEREBROVASCULAR DISEASE

Low mtDNA-CN was identified as a causal mediator of both dementia (Study 2; Chapter 4) and post-stroke outcomes (Study 3; Chapter 5) suggesting that interventions that increase mtDNA-CN levels may represent a novel therapeutic strategy for these conditions. As a corollary, drugs for mtDNA depletion disorders may be repurposed to prevent dementia and attenuate post-stroke outcomes. Although there are no clinically accepted treatments for mtDNA depletion, we will discuss emerging targets with pre-clinical support.

#### 6.4.3.2 NUCLEOSIDE-BASED SUPPLEMENTATION FOR TREATMENT OF MTDNA DEPLETION

Nucleoside supplementation has been shown to replenish mtDNA levels in animal models of mtDNA depletion in which genetic mutations cause deficiencies in nucleotide metabolism enzymes that are critical for mtDNA synthesis [22,23]. A phase I/II RCT (NCT03639701) is currently recruiting participants to test thymidine supplementation in

patients with TK2-deficient mtDNA depletion syndrome[24]. However, treatment efficacy is likely restricted to individuals with genetic defects in thymidine metabolism, since nucleoside supplementation was not found to alter mtDNA-CN levels in wildtype cells.

Another nucleotide-based compound with potentially broader application for treating mitochondrial dysfunction secondary to mtDNA depletion is nicotinamide adenine nucleotide (NAD)+. NAD+ is the well-known cofactor of the electron transport chain but also has other roles in metabolism, cell death, and immunity[25]. A drug library screen searching for compounds to rescue the ATP deficit caused by mtDNA depletion identified nicotinamide adenine nucleotide (NAD), the precursor to NAD+[26]. Specifically, NAD supplementation rescues ATP deficits in both *DGUOK* and *RRM2B* CRISPR/Cas9 knockout iSPC-derived hepatocytes. Furthermore, in-depth functional analyses in DGUOK-deficient hepatocytes revealed that administration of NAD rescues a broad range of mitochondrial defects including oxidative stress, mitochondrial membrane potential issues, and morphological aberrations. Also, oral administration of a NAD precursor, nicotinamide riboside (NR), to DGUOK-deficient rats promoted the expression of mtDNA-encoded electron transport genes and restored hepatic ATP levels *in vivo*. Beyond forms of mtDNA depletion characterized by nucleotide metabolism defects (i.e. *DGUOK*, *RRM2B*, and *TK2*), emerging evidence suggests that NAD+ upregulation may also be beneficial in the context of other forms caused by mutations in the core mtDNA replication machinery. For example, mitochondrial transcription factor A (TFAM) is responsible for initiating mtDNA replication and organizing mtDNA into nucleoids. Oller *et al.* (2021) found that mice with TFAM deficiency confined to vascular smooth muscle cells, develop aortic

aneurysms; NR supplementation is sufficient to derepress TFAM expression and mtDNA-CN levels thus reversing aortic aneurysm[27]. Altogether, these findings indicate that NR supplementation, a precursor to NAD and NAD+, as a potential treatment for mtDNA depletion irrespective of the underlying genetic mutation.

#### 6.4.3.3 EXPERIMENTAL EVIDENCE FOR NEUROPROTECTIVE EFFECTS OF NAD+ UPREGULATION

*In vitro* and *in vivo* rodent experiments suggest that upregulation of the NAD+ axis also protects against both ischemic and hemorrhagic stroke and that such neuroprotection, in part, is mediated by key mitochondrial regulators. First, just as mtDNA-CN levels are depleted after ischemic stroke, a similar drop in brain NAD occurs[25]. Exogenous supplementation of NAD+ restores DNA repair activity thereby mitigating ischemic cell death in oxygen and glucose-deprived rat neurons [25]. Second, nicotinamide phosphribosyltransferase (NAMPT), the rate-limiting enzyme in the production of NAD+, protects against cerebral injury and promotes neurogenesis after cerebral ischemia[28,29]. In rats with middle cerebral artery occlusion, pharmacological inhibition of NAMPT exacerbates neuronal cell death and increases the size of infarcts. Conversely, NAMPT overexpression overcomes these deficits by promoting neuronal survival via sirtuin-1-mediated induction of AMPK, an upstream mitochondrial biogenesis activator [28]. Third, oxidative stress and neuroinflammation following experimentally induced intracerebral hemorrhage in mice is rescued by acute administration (30-minutes post-stroke) of the NAD precursor, nicotinamide mononucleotide (NMN) [30]. NMN-treated mice also exhibit faster neurological recovery after intracerebral hemorrhage than those without treatment[30].

Notably, NMN treatment promoted the expression of nuclear respiratory factor 2, a gatekeeper of mitochondrial biogenesis[30].

#### 6.4.3.4 CLINICAL TRIALS OF NAD+ SUPPLEMENTATION

NAD precursor supplementation as a therapy for age-related diseases such as cardiometabolic disease, has garnered interest from academic and commercial sectors alike, culminating in several early phase RCTs. These trials suggest that daily NAD precursor supplementation (i) leads to persistent elevations in peripheral blood cell NAD+, (ii) is well-tolerated with no serious adverse events, and (iii) demonstrates preliminary evidence for protective vascular effects on blood pressure and arterial stiffness[31,32]. The next phase of RCTs is currently underway to test NAD precursor supplementation to ameliorate progression of various diseases including but not limited to rare mitochondrial disorders, dementia, mild cognitive decline, Parkinson's disease, and stroke. In fact, one trial (NCT03432871) is investigating NR supplementation for individuals with mitochondrial encephalopathy with lactic acidosis and stroke-like episodes [24]. To our knowledge, no trial is currently planned to investigate NR supplementation in a series of mtDNA depletion syndrome patients specifically, though this may be the next logical disease indication to explore.

## 6.5 STRENGTHS, LIMITATIONS, AND FUTURE DIRECTIONS

### 6.5.1 STRENGTHS

The main strengths of this work are: (i) the use of MR analysis, a statistical genetics framework robust to reverse causation and other types of confounding, to approximate causal effects of biomarkers on disease risk, (ii) the dual and context-appropriate

application of MR analysis to (a) screen circulating proteins with various biological functions for causal mediators and (b) to verify the role of an emerging marker of mitochondrial dysfunction with a strong biological prior to mediate stroke progression, (iii) the incorporation of public datasets to expand the scope of testable biomarkers and to boost statistical power to detect associations with disease, (iv) the sensitivity analyses and validation efforts that were employed to ensure the robustness of findings (e.g. proteomics technology-stratified MR analyses, benchmarking of AutoMitoC with qPCR-based measurements, etc.), and (v) the breadth of study designs employed to interrogate causal mediators of cerebrovascular disease (e.g. regression-based association testing, GWAS, ExWAS, MR, etc.).

### 6.5.2 LIMITATIONS

In Study 1 (Chapter 3), we implemented a methodological framework to conduct systematic proteome-wide MR screens for circulating protein mediators of stroke standing on the shoulders of colleagues who deployed similar methods for the discovery of putative causal mediators for heart disease, blood pressure, diabetes, and chronic kidney disease. We extended this approach to an agnostic investigation of phenotypes to forecast potential repurposing opportunities and adverse effects associated with target manipulation. Although at the time of analysis, this represented an improvement of our lab's existing pipelines, there are several improvements that could be made now due to methodological advances and new databases. First, colocalization analyses could be performed to provide assurance that the same genetic signals responsible for changing biomarker levels also accounts for changes in stroke susceptibility. Second, rather than assume that the effects of

causal biomarkers on disease act via circulation, the integration of tissue-specific eQTL data from the GTEx database may aid in pinpointing causal tissues. One of the challenges of interpreting biomarker MR analyses is tissue specificity. Because germline genetic variants persist in all cell types, the association between genetically determined circulating levels and disease outcomes may not be specific to blood. For example, the aforementioned SCARA5 cis-pQTL (rs2726927) that is associated with circulating SCARA5 levels and pro-thrombotic factors also influences *SCARA5* expression in the spleen, subcutaneous adipose tissue, and tibial nerve with concordant effect direction across all tissues. (https://www.gtexportal.org/home/gene/SCARA5). Therefore, it is difficult to disentangle the relevant and causal tissue through which SCARA5 activity protects against stroke.. Third, replication of secondary effects identified by the phenome-wide MR analysis can now be executed using newly available genome-wide and phenome-wide datasets, such as those made available by the FinnGen consortia. Fourth, the druggability or tractability of putative targets (e.g. whether there is a binding site or epitope on the target protein for small molecule or antibody binding) could be considered; however, the advent of oligonucleotide-based therapies that can down or up-regulate protein translation makes this consideration less constraining.

General limitations of using MR analysis for drug target evaluation include (i) potential for causal effects to act through alternative pathways apart from the target in question (i.e. horizontal pleiotropy), (ii) uncertain validity of extrapolating lifelong genetic effects for predicting the effects of pharmacological modulation, (iii) statistical power

varies by biomarker due to differences in instrument strength, and (iv) the inability to predict off-target side-effects of pharmacological interventions.

### 6.5.3 FUTURE DIRECTIONS FOR POTENTIAL THERAPEUTIC TARGETS

#### 6.5.3.1 SCARA5

Beyond independent replication of SCARA5 MR findings, there are several investigations that could be pursued to better gauge the therapeutic potential of SCARA5. First, there is a knowledge gap between our understanding of the membrane-bound vs. circulating form of SCARA5, and thus it would be of particular interest to know whether the circulating form has similar functions to its membrane-bound counterpart. For example, a key question is whether circulating SCARA5 is also capable of facilitating endocytosis of pro-thrombotic factors, and if not, then what are the ways in which SCARA5 protects against stroke? Second, elucidating the mechanisms through which circulating SCARA5 is generated may provide another angle for pharmacologic intervention. The generation of circulating SCARA5 presumably involves proteolytic cleavage of membrane-bound SCARA5, and notably, among the many predicted cleavage sites (https://web.expasy.org/cgi-bin/peptide_cutter/peptidecutter.pl) is a site for coagulation FXa, the pharmacological target of several anticoagulants (e.g. Rivaroxaban). Third, animal model experiments are warranted to directly test the effects of SCARA5 upregulation on stroke risk and key side-effects, such as bleeding and iron deficiency.

#### 6.5.3.2 TNFSF12

Although results suggest an unfavourable safety profile if TNFSF12 were to be directly targeted for stroke prevention, they still point towards a causal pathway potentially amenable to therapeutic targeting. Delving into downstream targets of TNFSF12 may provide a new source of targets with a more specific effect on stroke risk and favourable safety profile. This could be accomplished by performing a candidate MR analysis of all known downstream targets of TNFSF12 based on protein-protein interaction and pathway databases. Also, recent biomarker studies allude to TNFSF12 as a novel biomarker for post-intracerebral hemorrhage outcomes. Accordingly, subsequent MR analyses may be useful to help clarify whether these associations are causal, analogous to what was done for mtDNA-CN and post-ischemic stroke outcomes.

#### 6.5.3.3 MTDNA-CN RECOVERY FOR STROKE PROTECTION

In conjunction with animal model experiments, our findings suggest that restoring mtDNA-CN during the acute phase of stroke may be a novel therapeutic angle to attenuate post-stroke outcomes. Awaiting the results of preliminary NAD precursor trials, these will inform their safety in stroke patients specifically. Once confirmed, it may be conceivable to plan an efficacy trial investigating post-stroke NAD precursor supplementation to improve functional outcomes after stroke. To our knowledge, no trial has been proposed to test this specific hypothesis. Beyond clinical testing, additional experimental studies could be performed to further examine the relationship between cellular mtDNA content and existing neuroprotective mechanisms involving mitochondria. For example, although mtDNA depletion abolishes pro-survival signals normally elicited by horizontal mitochondrial transfer in cancer cell lines[33], there are no analogous investigations in oxygen

and glucose-deprived neurons or animal models. Moreover, the dependency of intercellular mitochondrial transfer on mtDNA levels has not been explored beyond the aforementioned study. This is a key missing link that could connect not only mtDNA regulation to post-stroke neuronal resiliency but also mtDNA regulation to cellular resilience in general.

#### 6.5.3.4 NEW HORIZONS FOR MTDNA-CN RESEARCH

Most epidemiological investigations (including the present work) have examined mtDNA content of aggregate immune cell populations from whole blood or buffy coat (Figure 6.1). As such, single-cell mtDNA-CN profiling will aid in pinpointing relevant cell populations that contribute to neuroprotection. Also, extracellular sources of circulating mtDNA-CN are relatively understudied but initial studies suggest a strong link with stroke pathophysiology. Extracellular mtDNA-CN consists of (i) free-floating and metabolically active mitochondria that may be horizontally transferred during diseases states to protect against injury, (ii) extracellular vesicles (microvesicles and exosomes) containing whole mtDNA genomes that are believed to participate in similar cytoprotective intercellular signalling mechanisms, (iii) and free-circulating mtDNA extruded from immune cells is characteristic of the damage associated molecular patterns that instigates IL1β-IL6-mediated inflammation[34]. Ultimately, deeper profiling of intracellular sources of mtDNA-CN will aid in clarifying the associations identified in the present work (Studies 1 and 2; Chapters 4 and 5), and comprehensive profiling of extracellular mtDNA-CN is an exciting potential source of new stroke biomarkers as they likely capture distinct mitochondrial processes relevant to stroke pathophysiology.

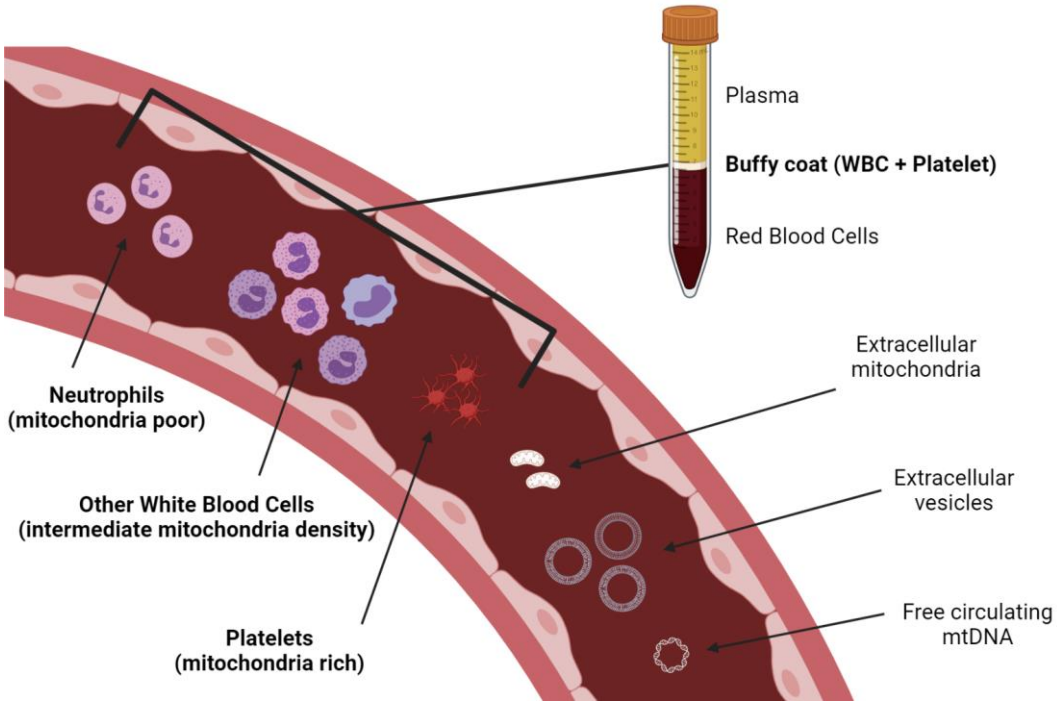



**Figure 6.1.** Sources of mtDNA-CN within whole blood. Created with https://biorender.com/.

## 6.6 CONCLUSION

Genomic, proteomic, and phenotypic datasets were integrated to elucidate putative circulating mediators of stroke risk and post-stroke outcomes. First, we conducted the largest MR screen of the circulating proteome for ischemic stroke mediators and identified two novel targets, TNFSF12 and SCARA5. Second, to understand the genetic determinants of an emerging mitochondrial biomarker in mtDNA-CN, we first devised a novel pipeline for array-based estimation tailored for large, multiethnic biobank studies. Third, using this pipeline, we carried out the largest and most comprehensive genetic association study for mtDNA-CN to date and increased the number of known loci by 40%. Fourth, we used a

subset of identified loci as genetic instruments to approximate genetically determined mtDNA-CN levels and provided the first MR evidence supporting causal roles for low mtDNA-CN in potentiating elevated risk for dementia and worse post-stroke outcomes.

## 6.7 REFERENCES


1. Robin, E. D. & Wong, R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J. Cell. Physiol.* **136**, 507–513 (1988).
2. Hägg, S., Jylhävä, J., Wang, Y., Czene, K. & Grassmann, F. Deciphering the genetic and epidemiological landscape of mitochondrial DNA abundance. *Hum. Genet.* (2020) doi:10.1007/s00439-020-02249-w.
3. Li, J. Y. *et al.* NIH Public Access. **16**, 35–46 (2010).
4. Ojala, J. R. M., Pikkarainen, T., Elmberger, G. & Tryggvason, K. Progressive reactive lymphoid connective tissue disease and development of autoantibodies in scavenger receptor A5-deficient mice. *Am. J. Pathol.* **182**, 1681–1695 (2013).
5. Yu, B. *et al.* Interactions of ferritin with scavenger receptor class A members. *J. Biol. Chem.* **295**, 15727–15741 (2020).
6. Yan, N. *et al.* Therapeutic upregulation of class a scavenger receptor member 5 inhibits tumor growth and metastasis. *Cancer Sci.* **103**, 1631–1639 (2012).
7. Mendes-Jorge, L. *et al.* L-ferritin binding to Scara5: A new iron traffic pathway potentially implicated in retinopathy. *PLoS One* **9**, 1–13 (2014).
8. WANG, J., WANG, S., CHEN, L. & TAN, J. SCARA5 suppresses the proliferation and migration, and promotes the apoptosis of human retinoblastoma

cells by inhibiting the PI3K/AKT pathway. *Mol. Med. Rep.* **23**, 1–10 (2021).

9. Lindström, S. *et al.* Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. *Blood* **134**, 1645–1657 (2019).
10. Swystun, L. L. *et al.* The scavenger receptor SCARA5 is an endocytic receptor for von Willebrand factor expressed by littoral cells in the human spleen. *J. Thromb. Haemost.* **17**, 1384–1396 (2019).
11. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
12. Smith, N. L. *et al.* Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von willebrand factor: The charge (cohorts for heart and aging research in genome epidemiology) consortium. *Circulation* **121**, 1382–1392 (2010).
13. Liu, Y. *et al.* Identification of SCARA5 as a Potential Biomarker for Oral Squamous Cell Carcinoma using MALDI–TOF–MS Analysis. *Proteomics - Clin. Appl.* **12**, 1–8 (2018).
14. Nerea, M., Guti, C., Bl, R., Mart, J. L. & Blanco-colio, L. M. Tumor Necrosis Factor-Like Weak Inducer of Apoptosis (TWEAK)/Fibroblast Growth Factor-Inducible 14 (Fn14) Axis in Cardiovascular Diseases: Progress and Challenges. *Cells* **14**, (2020).
15. da Silva-Candal, A. *et al.* The presence of leukoaraiosis enhances the association between sTWEAK and hemorrhagic transformation. *Ann. Clin. Transl. Neurol.* **7**,

2103–2114 (2020).

16. Da Silva-Candal, A. *et al.* sTWEAK is a marker of early haematoma growth and leukoaraiosis in intracerebral haemorrhage. *Stroke Vasc. Neurol.* 1–8 (2021) doi:10.1136/svn-2020-000684.

17. Wang, Q. *et al.* A phenome-wide multi-directional Mendelian randomization analysis of atrial fibrillation. *medRxiv* 1–22 (2020) doi:10.1101/2020.10.15.20212654.

18. Lam, E. T. *et al.* Phase I study of enavatuzumab, a first-in-class humanized monoclonal antibody targeting the TWEAK receptor, in patients with advanced solid tumors. *Mol. Cancer Ther.* **17**, 215–221 (2018).

19. R, F. *et al.* Evaluation of the Efficacy, Safety, and Tolerability of BIIB023 As an Adjunct to Standard of Care in Subjects with Lupus Nephritis. in *Arthritis Rheumatol.* (2016).

20. Galluppi, G. R., Wisniacki, N. & Stebbins, C. Population pharmacokinetic and pharmacodynamic analysis of BIIB023, an anti-TNF-like weak inducer of apoptosis (anti-TWEAK) monoclonal antibody. *Br. J. Clin. Pharmacol.* 118–128 (2016) doi:10.1111/bcp.12914.

21. Méndez-Barbero, N. *et al.* A major role of TWEAK/Fn14 axis as a therapeutic target for post-angioplasty restenosis. *EBioMedicine* **46**, 274–289 (2019).

22. Munro, B., Horvath, R. & Müller, J. S. Nucleoside supplementation modulates mitochondrial DNA copy number in the dguok-/- zebrafish. *Hum. Mol. Genet.* **28**, 796–803 (2019).

23. Vanden Avond, M. A. *et al.* The nucleotide prodrug CERC-913 improves mtDNA content in primary hepatocytes from DGUOK-deficient rats. *J. Inherit. Metab. Dis.* **44**, 492–501 (2021).

24. Almannai, M., El-hattab, A. W., Ali, M. & Soler-alfonso, C. Clinical trials in mitochondrial disorders, an update Mohammed. *Mol. Genet. Metab.* **131**, (2020).

25. Wang, S. *et al.* Cellular NAD replenishment confers marked neuroprotection against ischemic cell death: Role of enhanced DNA repair. *Stroke* **39**, 2587–2595 (2008).

26. Jing, R. *et al.* Potential Treatment for mtDNA Depletion Syndrome. *Cell Rep.* **25**, 1469–1484 (2018).

27. Oller, J. *et al.* Extracellular Tuning of Mitochondrial Respiration Leads to Aortic Aneurysm. *Circulation* (2021) doi:10.1161/circulationaha.120.051171.

28. Wang, P. *et al.* Nicotinamide phosphoribosyltransferase protects against ischemic stroke through SIRT1-dependent adenosine monophosphate-activated kinase pathway. *Ann. Neurol.* **69**, 360–374 (2011).

29. Zhao, Y. *et al.* Regenerative Neurogenesis after Ischemic Stroke Promoted by Nicotinamide Phosphoribosyltransferase-Nicotinamide Adenine Dinucleotide Cascade. *Stroke* **46**, 1966–1974 (2015).

30. Wei, C. C. *et al.* Nicotinamide mononucleotide attenuates brain injury after intracerebral hemorrhage by activating Nrf2/HO-1 signaling pathway. *Sci. Rep.* **7**, 1–13 (2017).

31. Dellinger, R. W. *et al.* Repeat dose NRPT (nicotinamide riboside and

pterostilbene) increases NAD+ levels in humans safely and sustainably: a randomized, double-blind, placebo-controlled study. *npj Aging Mech. Dis.* **3**, (2017).

32. Conze, D., Brenner, C. & Kruger, C. L. Safety and Metabolism of Long-term Administration of NIAGEN (Nicotinamide Riboside Chloride) in a Randomized, Double-Blind, Placebo-controlled Clinical Trial of Healthy Overweight Adults. *Sci. Rep.* **9**, 1–13 (2019).

33. Dong, L. F. *et al.* Horizontal transfer of whole mitochondria restores tumorigenic potential in mitochondrial DNA-deficient cancer cells. *Elife* **6**, 1–22 (2017).

34. Yamazoe, M. *et al.* Sparsely methylated mitochondrial cell free DNA released from cardiomyocytes contributes to systemic inflammatory response accompanied by atrial fibrillation. *Sci. Rep.* **11**, 1–10 (2021).

# APPENDIX A:

## Supplementary Data for Study 1

# Supplemental Tables

**Supplemental Table 1. Characteristics of individual biomarker GWAS used for genetic instrument selection**

| Reference | Study Cohort(s) | Cohort Description | GWAS Sample Size | Number of Biomarkers Analyzed | Biomarker Assay | Blood Fraction Tested |
|---|---|---|---|---|---|---|
| Ahola *et al.* (2017)[1] | YFS ; FINRISK | Population-based cohort studies | 2019;6323 (8332) | 41 | Bio-Rad Bio-Plex Pro Human Cytokine 27-plex and 21-plex assays | Serum ; Plasma |
| Folkersen *et al.* (2017)[2] | IMPROVE | Observational study of individuals with 3+ CVD risk factors but no symptoms | 3394 | 83 | O-link ProSeek CVD Array I | Plasma |
| Sjaarda *et al.* (2018)[3] | ORIGIN | Randomized controlled trial for secondary CVD prevention in diabetic and pre-diabetic individuals | 4147 | 227 | Myriad RBM Human Explorer Multi-Analyte Profile | Serum |
| Suhre *et al.* (2017)[4] | KORA F4; QMDiab | Population-based cohort study ; Cross-sectional case-control study of diabetes | 997;338 (1335) | 1124 | SOMAScan assay V3.2 | Plasma |
| Sun *et al.* (2017)[5] | INTERVAL | Randomized controlled trial of varying blood donation intervals in overall healthy individuals | 3301 | 2994 | SOMAScan assay (Extended Panel) | Plasma |
| **Total** | - | | **20509** | **3090*** | - | - |

*Number of unique biomarkers measured

**Supplemental Table 2. Genetic instruments used at each stage of the study for identified biomarkers**

| | SNPs | | |
|---|---|---|---|
| | **Stage 1** | **Stage 2** | **Stage 3** |
| **ABO** | rs116842520,rs139744322,rs144596390,rs886017,rs118070731,rs117338100,rs78338045,rs7027827,rs35637464,rs4962097,rs10901239,rs11793170,rs140167248,rs141808149,rs75218086,rs117329881,rs71503180,rs78755596,rs12216891,rs138374634,rs140167248,rs141808149,rs144596390,rs145834222,rs149181677,rs150258577,rs192308651,rs2073826,rs2285488,rs3124758,rs34008151,rs34021886,rs36218903,rs36221449,rs36221464,rs3761824,rs4129717,rs4962095,rs535015446,rs55988407,rs611983,rs61751475,rs633862,rs655911,rs66697526,rs7032208,rs7039497,rs71483206,rs71503180,rs73550898,rs78338045,rs8176704,rs9650778 | rs10901239,rs116842520,rs117338100,rs11793170,rs117965396,rs118070731,rs12216891,rs138374634,rs140167248,rs141808149,rs144596390,rs145834222,rs149181677,rs150258577,rs192308651,rs2073826,rs2285488,rs3124758,rs34008151,rs34021886,rs36218903,rs36221449,rs36221464,rs3761824,rs4129717,rs4962095,rs535015446,rs55988407,rs611983,rs61751475,rs633862,rs655911,rs66697526,rs7032208,rs7039497,rs71483206,rs71503180,rs73550898,rs78338045,rs8176704,rs9650778 | rs10901239,rs116842520,rs117338100,rs11793170,rs117965396,rs118070731,rs12216891,rs138374634,rs140167248,rs141808149,rs144596390,rs145834222,rs149181677,rs150258577,rs192308651,rs2073826,rs2285488,rs3124758,rs34008151,rs34021886,rs36218903,rs36221449,rs36221464,rs3761824,rs4129717,rs4962095,rs535015446,rs55988407,rs611983,rs61751475,rs633862,rs655911,rs66697526,rs7032208,rs7039497,rs71483206,rs71503180,rs73550898,rs78338045,rs8176704,rs9650778 |
| **CD40** | rs7267295,rs139518651,rs139824725,rs117871956,rs144858460,rs6073990,rs79522550,rs17344810,rs76659719,rs1009373,rs1569723,rs62215622,rs2024571,rs6074041,rs6124768,rs2425754,rs6065932,rs74495637,rs117606004,rs118016871,rs2425780,rs117967218,rs75838841 | rs1009373,rs117606004,rs117871956,rs117967218,rs118016871,rs139518651,rs139824725,rs144858460,rs1569723,rs17344810,rs2024571,rs2425754,rs2425780,rs6065932,rs6073990,rs6074041,rs6124768,rs62215622,rs7267295,rs74495637,rs75838841,rs76659719,rs79522550 | rs1009373,rs117606004,rs117871956,rs117967218,rs118016871,rs139518651,rs139824725,rs144858460,rs1569723,rs17344810,rs2024571,rs2425754,rs2425780,rs6065932,rs6073990,rs6074041,rs6124768,rs62215622,rs7267295,rs74495637,rs75838841,rs76659719,rs79522550 |
| **F11** | rs4862658,rs7687961,rs4253248,rs4253406,rs2289252,rs4862674 | rs2289252,rs4253248,rs4253406,rs4862674,rs7687961 | rs2289252,rs4253248,rs4253406,rs4862674,rs7687961 |
| **LPA** | rs62440365,rs146184004,rs3004079,rs140481741,rs9364552,rs138799654,rs78439586,rs189839402,rs2457574,rs112376176,rs13213129,rs12214416,rs4708871,rs7761293,rs55730499,rs41272078,rs73596816,rs80145669,rs117174672,rs6905073,rs41269876,rs2115868,rs186696265,rs783144,rs187158158,rs1819138,rs783147,rs1317026,rs4252193,rs10945689,rs139699952,rs75991907 | rs10945689,rs112376176,rs116881261,rs117174672,rs12214416,rs1317026,rs13213129,rs138799654,rs139699952,rs14048174146184004,rs1819138,rs186696265,rs189839402,rs2115868,rs2457574,rs3004079,rs34498812,rs41259144,rs41269876,rs41272078,rs41272114,rs4252193,rs4708871,rs55730499,rs62440365,rs6905073,rs73596816,rs75991907,rs7761293,rs783144,rs783147,rs78439586,rs80145669,rs9364552 | rs10945689,rs112376176,rs116881261,rs117174672,rs12214416,rs1317026,rs13213129,rs138799654,rs139699952,rs14048174146184004,rs1819138,rs186696265,rs189839402,rs2115868,rs2457574,rs3004079,rs34498812,rs41259144,rs41269876,rs41272078,rs41272114,rs4252193,rs4708871,rs55730499,rs62440365,rs6905073,rs73596816,rs75991907,rs7761293,rs783144,rs783147,rs78439586,rs80145669,rs9364552 |
| **MMP12** | rs112743271,rs2845675,rs35231465,rs11225397,rs3758861,rs118090914,rs1938896,rs17293152,rs470155,rs17359230,rs17860962,rs17878905,rs17881127,rs17885595,rs552306,rs17360292,rs112246582,rs114176245,rs613804,rs11225446,rs1892971,rs112759287,rs17860561,rs184540219,rs10791605,rs2155240,rs685286,rs11225507,rs151104286,rs11822520,rs117674362,rs186287714,rs4430479 | rs10791605,rs112246582,rs11225397,rs11225507,rs112743271,rs112759287,rs114176245,rs117674362,rs118090914,rs11822520,rs147031000,rs151104286,rs17293152,rs17359230,rs17360292,rs17860561,rs17860962,rs17878905,rs17881127,rs17885595,rs184540219,rs186287714,rs1892971,rs1938896,rs1940937,rs2155240,rs2845675,rs35231465,rs3758861,rs4430479,rs470155,rs471994,rs552306,rs570662,rs613804,rs685286 | rs10791605,rs112246582,rs11225397,rs11225507,rs112743271,rs112759287,rs114176245,rs117674362,rs118090914,rs11822520,rs147031000,rs151104286,rs17293152,rs17359230,rs17360292,rs17860561,rs17860962,rs17878905,rs17881127,rs17885595,rs184540219,rs186287714,rs1892971,rs1938896,rs1940937,rs2155240,rs2845675,rs35231465,rs3758861,rs4430479,rs470155,rs471994,rs552306,rs570662,rs613804,rs685286 |
| **SCARA5** | rs17384485,rs62498000,rs113456903,rs79124402,rs4732771,rs2726981,rs2685363,rs78282380,rs11994699,rs7831934,rs13268989,rs60047958,rs62496853,rs3757895 | rs113456903,rs11994699,rs13268989,rs17384485,rs2685363,rs2726981,rs3757895,rs4732771,rs60047958,rs62496853,rs62498000,rs78282380,rs7831934,rs79124402 | rs113456903,rs11994699,rs13268989,rs17384485,rs2685363,rs2726981,rs3757895,rs4732771,rs60047958,rs62496853,rs62498000,rs78282380,rs7831934,rs79124402 |
| **TNFSF12** | rs6503016,rs12937133,rs72842814,rs33989543,rs9906416,rs11078691,rs12948869,rs6608,rs150836621,rs1641511,rs1614984,rs2078486,rs11808193,rs3803802,rs12451505,rs56332843,rs117646332 | rs11078691,rs117646332,rs118081933,rs12451505,rs12948869,rs150836621,rs1614984,rs1641511,rs181975550,rs2078486,rs33989543,rs3803802,rs56206519,rs62062613,rs6608,rs72842814,rs9906416 | rs11078691,rs117646332,rs118081933,rs12451505,rs12948869,rs150836621,rs1614984,rs1641511,rs181975550,rs2078486,rs33989543,rs3803802,rs56206519,rs62062613,rs6608,rs72842814,rs9906416 |

SNP = single nucleotide polymorphism; ABO = Histo-blood group ABO system transferase; CD40 = cluster of differentiation 40; F11 = coagulation factor XI; LPA = apolipoprotein(a); MMP12 = matrix-metalloproteinase-12; SCARA5 = scavenger receptor class A5; TNFSF12 = tumour necrosis factor-like weak inducer of apoptosis.

**Supplemental Table 3. 679 disease traits analyzed in the Phe-MR analysis (Adapted from Zhou *et al.*, 2018)[6]**

| PheCode | Phenotype Description | Disease Category | Number of cases | Number of controls | Number of excluded controls |
|---|---|---|---|---|---|
| 401 | Hypertension | circulatory system | 77,977 | 330,366 | 618 |
| 401.1 | Essential hypertension | circulatory system | 77,723 | 330,366 | 872 |
| 550 | Abdominal hernia | digestive | 47,344 | 361,617 | 0 |
| 785 | Abdominal pain | symptoms | 41,316 | 367,645 | 0 |
| 716 | Other arthropathies | musculoskeletal | 38,715 | 365,819 | 4,427 |
| 716.9 | Arthropathy NOS | musculoskeletal | 37,043 | 365,819 | 6,099 |
| 272 | Disorders of lipid metabolism | endocrine/metabolic | 35,927 | 373,034 | 0 |
| 530 | Diseases of esophagus | digestive | 35,852 | 369,275 | 3,834 |
| 272.1 | Hyperlipidemia | endocrine/metabolic | 35,844 | 373,034 | 83 |
| 272.11 | Hypercholesterolemia | endocrine/metabolic | 33,242 | 373,034 | 2,685 |
| 530.1 | Esophagitis, GERD and related diseases | digestive | 32,108 | 369,275 | 7,578 |
| 418 | Nonspecific chest pain | circulatory system | 31,429 | 377,532 | 0 |
| 411 | Ischemic Heart Disease | circulatory system | 31,355 | 377,103 | 503 |
| 535 | Gastritis and duodenitis | digestive | 28,941 | 378,124 | 1,896 |
| 306 | Other mental disorder | mental disorders | 28,791 | 365,476 | 14,694 |
| 740 | Osteoarthrosis | musculoskeletal | 28,439 | 380,522 | 0 |
| 562 | Diverticulosis and diverticulitis | digestive | 27,311 | 334,783 | 46,867 |
| 562.1 | Diverticulosis | digestive | 27,268 | 334,783 | 46,910 |
| 550.2 | Diaphragmatic hernia | digestive | 27,126 | 361,617 | 20,218 |
| 495 | Asthma | respiratory | 26,332 | 375,505 | 7,124 |
| 427 | Cardiac dysrhythmias | circulatory system | 24,681 | 380,919 | 3,361 |
| 599 | Other symptoms/disorders or the urinary system | genitourinary | 24,031 | 384,930 | 0 |

Full table available here:
https://www.ahajournals.org/action/downloadSupplement?doi=10.1161%2FCIRCULATIONAHA.119.040180&file=circ_circulationaha-2019-040180_supp1.pdf

**Supplemental Table 4. List of 653 biomarkers examined in the primary MR screen for ischemic stroke mediators**

| Gene | Uniprot ID | Description |
|---|---|---|
| ABO | P16442 | Histo-blood group ABO system transferase (Fucosylglycoprotein 3-alpha-galactosyltransferase) (Fucosylglycoprotein alpha-N-acetylgalactosaminyltransferase) (Glycoprotein-fucosylgalactoside alpha-N-acetylgalactosaminyltransferase) (EC 2.4.1.40) (Glycoprotein-fucosylgalactoside alpha-galactosyltransferase) (EC 2.4.1.37) (Histo-blood group A transferase) (A transferase) (Histo-blood group B transferase) (B transferase) (NAGAT) [Cleaved into: Fucosylglycoprotein alpha-N-acetylgalactosaminyltransferase soluble form] |
| ACE | P12821 | Angiotensin-converting enzyme (ACE) (EC 3.2.1.-) (EC 3.4.15.1) (Dipeptidyl carboxypeptidase I) (Kininase II) (CD antigen CD143) [Cleaved into: Angiotensin-converting enzyme, soluble form] |
| ACP1 | P24666 | Low molecular weight phosphotyrosine protein phosphatase (LMW-PTP) (LMW-PTPase) (EC 3.1.3.48) (Adipocyte acid phosphatase) (Low molecular weight cytosolic acid phosphatase) (EC 3.1.3.2) (Red cell acid phosphatase 1) |
| ACP5 | P13686 | Tartrate-resistant acid phosphatase type 5 (TR-AP) (EC 3.1.3.2) (Tartrate-resistant acid ATPase) (TrATPase) (Type 5 acid phosphatase) |
| ADA2 | Q9NZK5 | Adenosine deaminase 2 (EC 3.5.4.4) (Cat eye syndrome critical region protein 1) |
| ADAM23 | O75077 | Disintegrin and metalloproteinase domain-containing protein 23 (ADAM 23) (Metalloproteinase-like, disintegrin-like, and cysteine-rich protein 3) (MDC-3) |
| ADAMTS13 | Q76LX8 | A disintegrin and metalloproteinase with thrombospondin motifs 13 (ADAM-TS 13) (ADAM-TS13) (ADAMTS-13) (EC 3.4.24.87) (von Willebrand factor-cleaving protease) (vWF-CP) (vWF-cleaving protease) |
| ADAMTS5 | Q9UNA0 | A disintegrin and metalloproteinase with thrombospondin motifs 5 (ADAM-TS 5) (ADAM-TS5) (ADAMTS-5) (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 11) (ADAM-TS 11) (ADAMTS-11) (ADMP-2) (Aggrecanase-2) |
| ADGRF5 | Q8IZF2 | Adhesion G protein-coupled receptor F5 (G-protein coupled receptor 116) |
| ADIPOQ | Q15848 | Adiponectin (30 kDa adipocyte complement-related protein) (Adipocyte complement-related 30 kDa protein) (ACRP30) (Adipocyte, C1q and collagen domain-containing protein) (Adipose most abundant gene transcript 1 protein) (apM-1) (Gelatin-binding protein) |
| AGT | P01019 | Angiotensinogen (Serpin A8) [Cleaved into: Angiotensin-1 (Angiotensin 1-10) (Angiotensin I) (Ang I); Angiotensin-2 (Angiotensin 1-8) (Angiotensin II) (Ang II); Angiotensin-3 (Angiotensin 2-8) (Angiotensin III) (Ang III) (Des-Asp[1]-angiotensin II); Angiotensin-4 (Angiotensin 3-8) (Angiotensin IV) (Ang IV); Angiotensin 1-9; Angiotensin 1-7; Angiotensin 1-5; Angiotensin 1-4] |
| AHSG | P02765 | Alpha-2-HS-glycoprotein (Alpha-2-Z-globulin) (Ba-alpha-2-glycoprotein) (Fetuin-A) [Cleaved into: Alpha-2-HS-glycoprotein chain A; Alpha-2-HS-glycoprotein chain B] |
| AKR1A1 | P14550 | Alcohol dehydrogenase [NADP(+)] (EC 1.1.1.2) (Aldehyde reductase) (Aldo-keto reductase family 1 member A1) |
| AKR1B1 | P15121 | Aldose reductase (AR) (EC 1.1.1.21) (Aldehyde reductase) (Aldo-keto reductase family 1 member B1) |

Full table available here:
https://www.ahajournals.org/action/downloadSupplement?doi=10.1161%2FCIRCULATIONAHA.119.040180&file=circ_circulationaha-2019-040180_supp1.pdf

**Supplemental Table 5. Sensitivity MR analyses for ischemic stroke subtype analyses**

| Biomarker | Stroke Subtype | MR Method | nsnp | OR | 95% CI | P-value | Biomarker r2 | Correct Causal Direction? | Steiger P-value | Egger Intercept P-value |
|---|---|---|---|---|---|---|---|---|---|---|
| ABO | CES | MR-EGGER | 39 | 1.05 | 1.00-1.10 | 0.045715 | 0.30 | TRUE | 0 | 0.68 |
| ABO | CES | IVW | 39 | 1.06 | 1.03-1.09 | 1.89E-05 | 0.30 | TRUE | 0 | |
| ABO | CES | MR-RAPS | 39 | 1.06 | 1.04-1.09 | 4.54E-07 | 0.30 | TRUE | 0 | |
| ABO | CES | WM | 39 | 1.06 | 1.02-1.10 | 0.001534 | 0.30 | TRUE | 0 | |
| ABO | LAA | MR-EGGER | 39 | 1.08 | 1.02-1.14 | 0.011521 | 0.30 | TRUE | 0 | 0.73 |
| ABO | LAA | IVW | 39 | 1.09 | 1.05-1.12 | 1.86E-07 | 0.30 | TRUE | 0 | |
| ABO | LAA | MR-RAPS | 39 | 1.08 | 1.05-1.12 | 2.43E-07 | 0.30 | TRUE | 0 | |
| ABO | LAA | WM | 39 | 1.07 | 1.03-1.12 | 0.000819 | 0.30 | TRUE | 0 | |
| F11 | CES | MR-EGGER | 6 | 1.15 | 0.97-1.37 | 0.197397 | 0.01 | TRUE | 0 | 0.55 |
| F11 | CES | IVW | 6 | 1.25 | 1.17-1.34 | 7.80E-11 | 0.01 | TRUE | 0 | |
| F11 | CES | MR-RAPS | 6 | 1.25 | 1.16-1.36 | 1.34E-08 | 0.01 | TRUE | 0 | |
| F11 | CES | WM | 6 | 1.24 | 1.14-1.34 | 2.32E-07 | 0.01 | TRUE | 0 | |
| SCARA5 | CES | MR-EGGER | 14 | 0.83 | 0.66-1.06 | 0.163432 | 0.01 | TRUE | 0 | 0.69 |
| SCARA5 | CES | IVW | 14 | 0.80 | 0.72-0.88 | 6.73E-06 | 0.01 | TRUE | 0 | |
| SCARA5 | CES | MR-RAPS | 14 | 0.78 | 0.70-0.88 | 1.46E-05 | 0.01 | TRUE | 0 | |
| SCARA5 | CES | WM | 14 | 0.79 | 0.69-0.90 | 0.000642 | 0.01 | TRUE | 0 | |
| TNFSF12 | CES | MR-EGGER | 17 | 0.81 | 0.71-0.92 | 0.006758 | 0.05 | TRUE | 0 | 0.33 |
| TNFSF12 | CES | IVW | 17 | 0.86 | 0.81-0.91 | 3.04E-07 | 0.05 | TRUE | 0 | |
| TNFSF12 | CES | MR-RAPS | 17 | 0.86 | 0.81-0.91 | 7.69E-07 | 0.05 | TRUE | 0 | |
| TNFSF12 | CES | WM | 17 | 0.84 | 0.78-0.91 | 1.15E-05 | 0.05 | TRUE | 0 | |
| CD40 | LAA | MR-EGGER | 23 | 0.80 | 0.64-1.00 | 0.058122 | 0.06 | TRUE | 0 | 0.34 |
| CD40 | LAA | IVW | 23 | 0.73 | 0.65-0.81 | 1.08E-08 | 0.06 | TRUE | 0 | |
| CD40 | LAA | MR-RAPS | 23 | 0.73 | 0.66-0.80 | 1.90E-10 | 0.06 | TRUE | 0 | |
| CD40 | LAA | WM | 23 | 0.72 | 0.63-0.82 | 1.68E-06 | 0.06 | TRUE | 0 | |
| LPA | LAA | MR-EGGER | 31 | 1.20 | 1.08-1.33 | 0.0016 | 0.12 | TRUE | 0 | 0.69 |
| LPA | LAA | IVW | 31 | 1.18 | 1.11-1.26 | 6.19E-07 | 0.12 | TRUE | 0 | |
| LPA | LAA | MR-RAPS | 31 | 1.22 | 1.14-1.30 | 3.19E-09 | 0.12 | TRUE | 0 | |
| LPA | LAA | WM | 31 | 1.24 | 1.15-1.33 | 1.40E-08 | 0.12 | TRUE | 0 | |
| MMP12 | LAA | MR-EGGER | 33 | 0.83 | 0.73-0.95 | 0.008978 | 0.14 | TRUE | 0 | 0.98 |
| MMP12 | LAA | IVW | 33 | 0.83 | 0.77-0.90 | 1.80E-06 | 0.14 | TRUE | 0 | |
| MMP12 | LAA | MR-RAPS | 33 | 0.83 | 0.77-0.90 | 6.56E-06 | 0.14 | TRUE | 0 | |
| MMP12 | LAA | WM | 33 | 0.80 | 0.74-0.87 | 9.52E-08 | 0.14 | TRUE | 0 | |

Effects are expressed per 1 SD increase in biomarker levels. MR-EGGER = MR-EGGER; IVW = Inverse Variance Weighted; MR-RAPS = MR-Robust adjusted profile score; WM = Weighted Median; ABO = Histo-blood group ABO system transferase; CD40 = cluster of differentiation 40; F11 = coagulation factor XI; LPA = apolipoprotein(a); MMP12 = matrix-metalloproteinase-12; SCARA5 = scavenger receptor class A5; TNFSF12 = tumour necrosis factor-like weak inducer of apoptosis; OR = odds ratio; CI = confidence interval.

**Supplemental Table 6. Association between stroke biomarkers and ischemic subtypes and total ischemic stroke**

| Biomarker | Stroke Subtype | # SNPs | Biomarker r2 | OR | 95% CI | P-value |
|---|---|---|---|---|---|---|
| ABO** | LAA | 39 | 0.3 | 1.08 | 1.05-1.12 | $2.43 \times 10^{-7}$ |
| ABO** | CES | 39 | 0.3 | 1.06 | 1.04-1.09 | $4.54 \times 10^{-7}$ |
| ABO | SAO | 39 | 0.3 | 1.01 | 0.98-1.04 | 0.47 |
| ABO** | Ischemic Stroke | 39 | 0.3 | 1.03 | 1.02-1.05 | $9.79 \times 10^{-8}$ |
| ABO* | All Stroke | 39 | 0.3 | 1.03 | 1.01-1.04 | $3.73 \times 10^{-5}$ |
| CD40** | LAA | 23 | 0.06 | 0.73 | 0.66-0.80 | $1.90 \times 10^{-10}$ |
| CD40 | CES | 23 | 0.06 | 1.05 | 0.97-1.13 | 0.21 |
| CD40* | SAO | 23 | 0.06 | 0.9 | 0.83-0.99 | 0.03 |
| CD40** | Ischemic Stroke | 23 | 0.06 | 0.91 | 0.87-0.95 | $3.29 \times 10^{-6}$ |
| CD40** | All Stroke | 23 | 0.06 | 0.91 | 0.87-0.95 | $9.07 \times 10^{-6}$ |
| F11 | LAA | 6 | 0.01 | 0.98 | 0.89-1.07 | 0.64 |
| F11** | CES | 6 | 0.01 | 1.25 | 1.16-1.36 | $1.34 \times 10^{-8}$ |
| F11 | SAO | 6 | 0.01 | 1.07 | 0.98-1.17 | 0.11 |
| F11** | Ischemic Stroke | 6 | 0.01 | 1.09 | 1.05-1.13 | $2.32 \times 10^{-5}$ |
| F11* | All Stroke | 6 | 0.01 | 1.07 | 1.03-1.10 | $2.49 \times 10^{-4}$ |
| LPA** | LAA | 31 | 0.12 | 1.22 | 1.14-1.30 | $3.19 \times 10^{-9}$ |
| LPA | CES | 31 | 0.12 | 1.02 | 0.98-1.06 | 0.4 |
| LPA | SAO | 31 | 0.12 | 0.98 | 0.92-1.03 | 0.41 |
| LPA* | Ischemic Stroke | 31 | 0.12 | 1.02 | 1.003-1.04 | 0.02 |
| LPA | All Stroke | 31 | 0.12 | 1.02 | 0.999-1.04 | 0.06 |
| MMP12** | LAA | 33 | 0.14 | 0.83 | 0.77-0.90 | $6.56 \times 10^{-6}$ |
| MMP12* | CES | 33 | 0.14 | 0.94 | 0.90-0.99 | 0.01 |
| MMP12* | SAO | 33 | 0.14 | 0.93 | 0.89-0.99 | 0.01 |
| MMP12** | Ischemic Stroke | 33 | 0.14 | 0.91 | 0.89-0.95 | $5.91 \times 10^{-13}$ |
| MMP12** | All Stroke | 33 | 0.14 | 0.93 | 0.91-0.95 | $9.16 \times 10^{-12}$ |
| SCARA5 | LAA | 14 | 0.02 | 1.08 | 0.90-1.28 | 0.41 |
| SCARA5** | CES | 14 | 0.02 | 0.78 | 0.70-0.88 | $1.46 \times 10^{-5}$ |
| SCARA5* | SAO | 14 | 0.02 | 0.86 | 0.76-0.98 | 0.02 |
| SCARA5* | Ischemic Stroke | 14 | 0.02 | 0.89 | 0.84-0.95 | $2.66 \times 10^{-4}$ |
| SCARA5* | All Stroke | 14 | 0.02 | 0.9 | 0.86-0.95 | $8.26 \times 10^{-5}$ |
| TNFSF12 | LAA | 17 | 0.05 | 0.99 | 0.92-1.07 | 0.8 |
| TNFSF12** | CES | 17 | 0.05 | 0.86 | 0.81-0.91 | $7.69 \times 10^{-7}$ |
| TNFSF12 | SAO | 17 | 0.05 | 1.04 | 0.97-1.12 | 0.24 |
| TNFSF12* | Ischemic Stroke | 17 | 0.05 | 0.96 | 0.93-0.99 | 0.01 |
| TNFSF12* | All Stroke | 17 | 0.05 | 0.94 | 0.84-1.05 | 0.3 |

* Nominally significant (P<0.05) ** Bonferroni significant ($P<2.55 \times 10^{-5}$). ABO = Histo-blood group ABO system transferase; CD40 = cluster of differentiation 40; F11 = coagulation factor XI; LPA = apolipoprotein(a); MMP12 = matrix-metalloproteinase-12; SCARA5 = scavenger receptor class A5; TNFSF12 = tumour necrosis factor-like weak inducer of apoptosis; LAA = large artery atherosclerosis; CES = cardioembolic stroke; SAO = small artery occlusion; OR = odds ratio; CI = confidence interval.

**Supplemental Table 7. Sensitivity MR analyses for hemorrhagic stroke subtype analyses.**

| Biomarker | Stroke Subtype | MR Method | nsnp | OR | 95 % CI | P-value | Biomarker r2 | Correct Causal Direction ? | Steiger P-value | Egger Intercept P-value |
|---|---|---|---|---|---|---|---|---|---|---|
| ABO | SAH | MR-EGGER | 41 | 0.99 | 0.90-1.08 | 0.77354 | 0.292 | TRUE | 0 | 0.79 |
| ABO | SAH | IVW | 41 | 1.00 | 0.95-1.05 | 0.90825 | 0.292 | TRUE | 0 | |
| ABO | SAH | MR-RAPS | 41 | 0.99 | 0.94-1.04 | 0.75015 | 0.292 | TRUE | 0 | |
| ABO | SAH | WM | 41 | 0.97 | 0.90-1.03 | 0.31904 | 0.292 | TRUE | 0 | |
| ABO | ICH | MR-EGGER | 41 | 0.92 | 0.82-1.03 | 0.1453 | 0.292 | TRUE | 0 | 0.13 |
| ABO | ICH | IVW | 41 | 0.99 | 0.93-1.06 | 0.72437 | 0.292 | TRUE | 0 | |
| ABO | ICH | MR-RAPS | 41 | 0.99 | 0.93-1.05 | 0.64274 | 0.292 | TRUE | 0 | |
| ABO | ICH | WM | 41 | 0.97 | 0.90-1.05 | 0.47618 | 0.292 | TRUE | 0 | |
| CD40 | SAH | MR-EGGER | 23 | 1.20 | 0.82-1.76 | 0.36341 | 0.064 | TRUE | 0 | 0.10 |
| CD40 | SAH | IVW | 23 | 0.89 | 0.73-1.09 | 0.26792 | 0.064 | TRUE | 0 | |
| CD40 | SAH | MR-RAPS | 23 | 0.88 | 0.74-1.04 | 0.12902 | 0.064 | TRUE | 0 | |
| CD40 | SAH | WM | 23 | 0.84 | 0.66-1.06 | 0.14999 | 0.064 | TRUE | 0 | |
| CD40 | ICH | MR-EGGER | 23 | 1.33 | 0.92-1.92 | 0.14014 | 0.064 | TRUE | 0 | 0.25 |
| CD40 | ICH | IVW | 23 | 1.10 | 0.92-1.32 | 0.30199 | 0.064 | TRUE | 0 | |
| CD40 | ICH | MR-RAPS | 23 | 1.09 | 0.89-1.34 | 0.38562 | 0.064 | TRUE | 0 | |
| CD40 | ICH | WM | 23 | 1.04 | 0.81-1.35 | 0.7487 | 0.064 | TRUE | 0 | |
| F11 | SAH | MR-EGGER | 5 | 0.80 | 0.51-1.27 | 0.41666 | 0.011 | TRUE | 0 | 0.43 |
| F11 | SAH | IVW | 5 | 0.98 | 0.82-1.17 | 0.79922 | 0.011 | TRUE | 0 | |
| F11 | SAH | MR-RAPS | 5 | 0.96 | 0.82-1.13 | 0.6047 | 0.011 | TRUE | 0 | |
| F11 | SAH | WM | 5 | 0.93 | 0.78-1.10 | 0.38641 | 0.011 | TRUE | 0 | |
| F11 | ICH | MR-EGGER | 5 | 0.87 | 0.56-1.36 | 0.58545 | 0.011 | TRUE | 0 | 0.83 |
| F11 | ICH | IVW | 5 | 0.92 | 0.77-1.09 | 0.30886 | 0.011 | TRUE | 0 | |
| F11 | ICH | MR-RAPS | 5 | 0.91 | 0.76-1.09 | 0.3242 | 0.011 | TRUE | 0 | |
| F11 | ICH | WM | 5 | 0.90 | 0.74-1.10 | 0.29924 | 0.011 | TRUE | 0 | |
| LPA | SAH | MR-EGGER | 35 | 1.06 | 0.95-1.18 | 0.29925 | 0.132 | TRUE | 0 | 0.34 |
| LPA | SAH | IVW | 35 | 1.10 | 1.03-1.18 | 0.0061 | 0.132 | TRUE | 0 | |
| LPA | SAH | MR-RAPS | 35 | 1.10 | 1.02-1.18 | 0.01038 | 0.132 | TRUE | 0 | |
| LPA | SAH | WM | 35 | 1.08 | 0.97-1.20 | 0.14282 | 0.132 | TRUE | 0 | |
| LPA | ICH | MR-EGGER | 35 | 1.01 | 0.88-1.15 | 0.90428 | 0.132 | TRUE | 0 | 0.74 |
| LPA | ICH | IVW | 35 | 0.99 | 0.91-1.08 | 0.84207 | 0.132 | TRUE | 0 | |
| LPA | ICH | MR-RAPS | 35 | 1.00 | 0.91-1.09 | 0.94617 | 0.132 | TRUE | 0 | |
| LPA | ICH | WM | 35 | 1.05 | 0.94-1.18 | 0.37634 | 0.132 | TRUE | 0 | |
| MMP12 | SAH | MR-EGGER | 36 | 1.04 | 0.88-1.23 | 0.63925 | 0.146 | TRUE | 0 | 0.70 |
| MMP12 | SAH | IVW | 36 | 1.07 | 0.97-1.17 | 0.16097 | 0.146 | TRUE | 0 | |
| MMP12 | SAH | MR-RAPS | 36 | 1.07 | 0.97-1.17 | 0.19779 | 0.146 | TRUE | 0 | |
| MMP12 | SAH | WM | 36 | 1.08 | 0.94-1.24 | 0.29873 | 0.146 | TRUE | 0 | |
| MMP12 | ICH | MR-EGGER | 36 | 1.22 | 0.99-1.50 | 0.07723 | 0.146 | TRUE | 0 | 0.99 |
| MMP12 | ICH | IVW | 36 | 1.21 | 1.08-1.37 | 0.00121 | 0.146 | TRUE | 0 | |
| MMP12 | ICH | MR-RAPS | 36 | 1.22 | 1.09-1.36 | 0.00048 | 0.146 | TRUE | 0 | |
| MMP12 | ICH | WM | 36 | 1.15 | 0.98-1.35 | 0.09762 | 0.146 | TRUE | 0 | |
| SCARA5 | SAH | MR-EGGER | 14 | 0.48 | 0.27-0.83 | 0.02186 | 0.015 | TRUE | 0 | 0.19 |
| SCARA5 | SAH | IVW | 14 | 0.68 | 0.53-0.86 | 0.00141 | 0.015 | TRUE | 0 | |
| SCARA5 | SAH | MR-RAPS | 14 | 0.67 | 0.52-0.85 | 0.0012 | 0.015 | TRUE | 0 | |
| SCARA5 | SAH | WM | 14 | 0.72 | 0.51-1.00 | 0.04801 | 0.015 | TRUE | 0 | |
| SCARA5 | ICH | MR-EGGER | 14 | 0.57 | 0.32-1.03 | 0.08604 | 0.015 | TRUE | 0 | 0.31 |
| SCARA5 | ICH | IVW | 14 | 0.76 | 0.60-0.97 | 0.03047 | 0.015 | TRUE | 0 | |
| SCARA5 | ICH | MR-RAPS | 14 | 0.77 | 0.59-1.00 | 0.04834 | 0.015 | TRUE | 0 | |
| SCARA5 | ICH | WM | 14 | 0.77 | 0.54-1.11 | 0.16059 | 0.015 | TRUE | 0 | |
| TNFSF12 | SAH | MR-EGGER | 17 | 1.23 | 0.92-1.66 | 0.18251 | 0.047 | TRUE | 0 | 0.47 |
| TNFSF12 | SAH | IVW | 17 | 1.36 | 1.20-1.55 | 1.13E-06 | 0.047 | TRUE | 0 | |
| TNFSF12 | SAH | MR-RAPS | 17 | 1.37 | 1.20-1.56 | 3.12E-06 | 0.047 | TRUE | 0 | |
| TNFSF12 | SAH | WM | 17 | 1.39 | 1.18-1.65 | 8.72E-05 | 0.047 | TRUE | 0 | |
| TNFSF12 | ICH | MR-EGGER | 17 | 1.27 | 0.91-1.77 | 0.17539 | 0.047 | TRUE | 0 | 0.65 |
| TNFSF12 | ICH | IVW | 17 | 1.37 | 1.19-1.57 | 1.4E-05 | 0.047 | TRUE | 0 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| TNFSF12 | ICH | MR-RAPS | 17 | 1.37 | 1.18-1.59 | 2.8E-05 | 0.047 | TRUE | 0 |
| TNFSF12 | ICH | WM | 17 | 1.34 | 1.11-1.63 | 0.00295 | 0.047 | TRUE | 0 |

Effects are expressed per 1 SD increase in biomarker levels. MR-EGGER = MR-EGGER; IVW = Inverse Variance Weighted; MR-RAPS = MR-Robust adjusted profile score; WM = Weighted Median; ABO = Histo-blood group ABO system transferase; CD40 = cluster of differentiation 40; F11 = coagulation factor XI; LPA = apolipoprotein(a); MMP12 = matrix-metalloproteinase-12; SCARA5 = scavenger receptor class A5; TNFSF12 = tumour necrosis factor-like weak inducer of apoptosis; SAH = subarachnoid hemorrhage; ICH = intracerebral hemorrhage; OR = odds ratio; CI = confidence interval.

**Supplemental Table 8. Association between identified biomarkers and ICH subtypes based on Woo et al. (2014)[7].**

| Biomarker | Stroke Subtype | Biomarker r2 | OR | 95% CI | P-value |
|---|---|---|---|---|---|
| ABO | non-lobar | 0.24 | 0.93 | 0.86-1.00 | 0.06 |
| ABO | lobar | 0.23 | 1.03 | 0.94-1.12 | 0.57 |
| CD40 | non-lobar | 0.05 | 1.01 | 0.75-1.37 | 0.95 |
| CD40 | lobar | 0.05 | 0.90 | 0.68-1.20 | 0.49 |
| F11 | non-lobar | 0.01 | 0.99 | 0.79-1.25 | 0.96 |
| F11 | lobar | 0.01 | 1.14 | 0.88-1.47 | 0.32 |
| LPA | non-lobar | 0.05 | 1.10 | 0.90-1.34 | 0.36 |
| LPA | lobar | 0.05 | 0.94 | 0.75-1.18 | 0.61 |
| MMP12 | non-lobar | 0.10 | 0.84 | 0.71-1.00 | 0.05 |
| MMP12 | lobar | 0.11 | 0.82 | 0.69-0.98 | 0.03 |
| SCARA5 | non-lobar | 0.01 | 0.92 | 0.65-1.30 | 0.63 |
| SCARA5 | lobar | 0.01 | 0.87 | 0.55-1.37 | 0.54 |
| TNFSF12 | non-lobar | 0.03 | 1.07 | 0.83-1.37 | 0.60 |
| TNFSF12 | lobar | 0.02 | 1.12 | 0.61-2.07 | 0.71 |

ABO = Histo-blood group ABO system transferase; CD40 = cluster of differentiation 40; F11 = coagulation factor XI; LPA = apolipoprotein(a); MMP12 = matrix-metalloproteinase-12; SCARA5 = scavenger receptor class A5; TNFSF12 = tumour necrosis factor-like weak inducer of apoptosis; CI = confidence interval.

**Supplemental Table 9. Sensitivity MR analyses for all significant Phe-WAS results.**

| Biomarker | PheCode | Phenotype Description | Disease Chapter | MR Method | nsnp | OR | 95% CI | P-value | Biomarker r2 | Egger Intercept P-value |
|---|---|---|---|---|---|---|---|---|---|---|
| ABO | 157 | Pancreatic cancer | neoplasms | MR-EGGER | 42 | 1.24 | 1.09-1.40 | 0.00215 | 0.303 | 0.80 |
| ABO | 157 | Pancreatic cancer | neoplasms | IVW | 42 | 1.22 | 1.14-1.31 | 5.87E-08 | 0.303 | 0.80 |
| ABO | 157 | Pancreatic cancer | neoplasms | MR-RAPS | 42 | 1.22 | 1.12-1.32 | 2.48E-06 | 0.303 | 0.80 |
| ABO | 157 | Pancreatic cancer | neoplasms | WM | 42 | 1.20 | 1.09-1.33 | 0.00024 | 0.303 | 0.80 |
| ABO | 174 | Breast cancer | neoplasms | MR-EGGER | 42 | 1.03 | 1.00-1.06 | 0.02724 | 0.303 | 0.49 |
| ABO | 174 | Breast cancer | neoplasms | IVW | 42 | 1.04 | 1.02-1.05 | 6.04E-07 | 0.303 | 0.49 |
| ABO | 174 | Breast cancer | neoplasms | MR-RAPS | 42 | 1.04 | 1.02-1.05 | 1.14E-06 | 0.303 | 0.49 |
| ABO | 174 | Breast cancer | neoplasms | WM | 42 | 1.04 | 1.02-1.06 | 0.00052 | 0.303 | 0.49 |
| CD40 | 202 | Cancer of other lymphoid, histiocytic tissue | neoplasms | MR-EGGER | 23 | 0.66 | 0.51-0.85 | 0.00393 | 0.064 | 0.29 |
| CD40 | 202 | Cancer of other lymphoid, histiocytic tissue | neoplasms | IVW | 23 | 0.74 | 0.65-0.84 | 3.39E-06 | 0.064 | 0.29 |
| CD40 | 202 | Cancer of other lymphoid, histiocytic tissue | neoplasms | MR-RAPS | 23 | 0.74 | 0.65-0.84 | 3.18E-06 | 0.064 | 0.29 |
| CD40 | 202 | Cancer of other lymphoid, histiocytic tissue | neoplasms | WM | 23 | 0.74 | 0.64-0.87 | 0.00022 | 0.064 | 0.29 |
| CD40 | 202.2 | Non-Hodgkins lymphoma | neoplasms | MR-EGGER | 23 | 0.68 | 0.50-0.93 | 0.02269 | 0.064 | 0.75 |
| CD40 | 202.2 | Non-Hodgkins lymphoma | neoplasms | IVW | 23 | 0.71 | 0.62-0.83 | 6.83E-06 | 0.064 | 0.75 |
| CD40 | 202.2 | Non-Hodgkins lymphoma | neoplasms | MR-RAPS | 23 | 0.70 | 0.60-0.82 | 7.93E-06 | 0.064 | 0.75 |
| CD40 | 202.2 | Non-Hodgkins lymphoma | neoplasms | WM | 23 | 0.73 | 0.61-0.87 | 0.00051 | 0.064 | 0.75 |
| ABO | 214 | Lipoma | neoplasms | MR-EGGER | 42 | 1.05 | 1.01-1.09 | 0.00952 | 0.303 | 0.92 |
| ABO | 214 | Lipoma | neoplasms | IVW | 42 | 1.05 | 1.03-1.07 | 4.11E-06 | 0.303 | 0.92 |
| ABO | 214 | Lipoma | neoplasms | MR-RAPS | 42 | 1.05 | 1.03-1.07 | 5.74E-06 | 0.303 | 0.92 |
| ABO | 214 | Lipoma | neoplasms | WM | 42 | 1.06 | 1.03-1.09 | 4.47E-05 | 0.303 | 0.92 |
| ABO | 214.1 | Lipoma of skin and subcutaneous tissue | neoplasms | MR-EGGER | 42 | 1.08 | 1.04-1.13 | 0.00049 | 0.303 | 0.31 |
| ABO | 214.1 | Lipoma of skin and subcutaneous tissue | neoplasms | IVW | 42 | 1.06 | 1.04-1.09 | 2.79E-07 | 0.303 | 0.31 |
| ABO | 214.1 | Lipoma of skin and subcutaneous tissue | neoplasms | MR-RAPS | 42 | 1.07 | 1.04-1.09 | 2.15E-07 | 0.303 | 0.31 |

Effects are expressed per 1 SD increase in biomarker levels. MR-EGGER = MR-EGGER; IVW = Inverse Variance Weighted; MR-RAPS = MR-Robust adjusted profile score; WM = Weighted Median; ABO = Histo-blood group ABO system transferase; CD40 = cluster of differentiation 40; F11 = coagulation factor XI; LPA = apolipoprotein(a); MMP12 = matrix-metalloproteinase-12; SCARA5 = scavenger receptor class A5; TNFSF12 = tumour necrosis factor-like weak inducer of apoptosis.

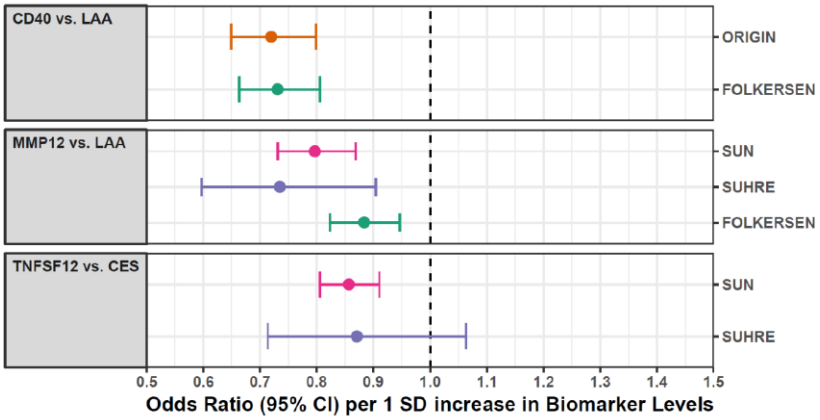
Full table available here:
https://www.ahajournals.org/action/downloadSupplement?doi=10.1161%2FCIRCULATIONAHA.119.040180&file=circ_circulationaha-2019-040180_supp1.pdf

**Supplemental Table 10. Association between select biomarkers serving as negative controls with ischemic stroke subtypes.**

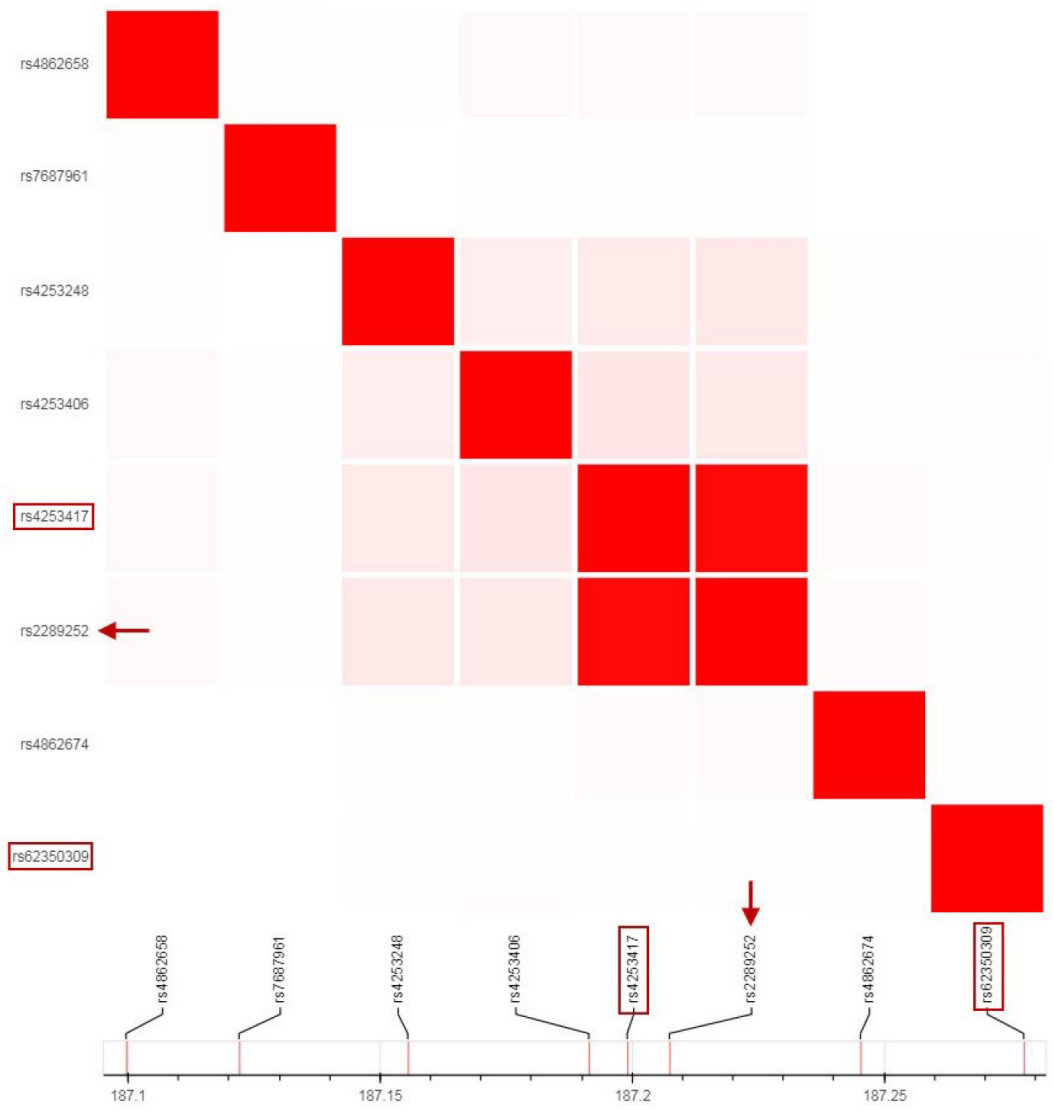
| Biomarker | Stroke Subtype | Biomarker r2 | OR | 95% CI | P-value |
|---|---|---|---|---|---|
| CHI3L1 | CES | 0.15 | 0.95 | 0.90-1.00 | 0.06 |
| CHI3L1 | LAA | 0.15 | 1.04 | 0.96-1.13 | 0.34 |
| CHI3L1 | SAO | 0.15 | 1.06 | 0.99-1.13 | 0.08 |
| CRP | CES | 0.019 | 0.99 | 0.84-1.15 | 0.86 |
| CRP | LAA | 0.019 | 1.17 | 0.95-1.45 | 0.14 |
| CRP | SAO | 0.019 | 1.03 | 0.86-1.22 | 0.78 |
| CST3 | CES | 0.015 | 0.97 | 0.85-1.09 | 0.59 |
| CST3 | LAA | 0.015 | 1.05 | 0.89-1.25 | 0.56 |
| CST3 | SAO | 0.015 | 0.96 | 0.82-1.12 | 0.58 |

YKL-40 (CHI3L1), Cystatin C (CST3), and C-reactive protein (CRP) have been epidemiologically associated with risk of stroke but subsequent MRs did not reveal causal associations[8–10]. CES = cardioembolic stroke; LAA = large artery atherosclerosis; SAO = small artery occlusion; OR = odds ratio; CI = confidence interval.

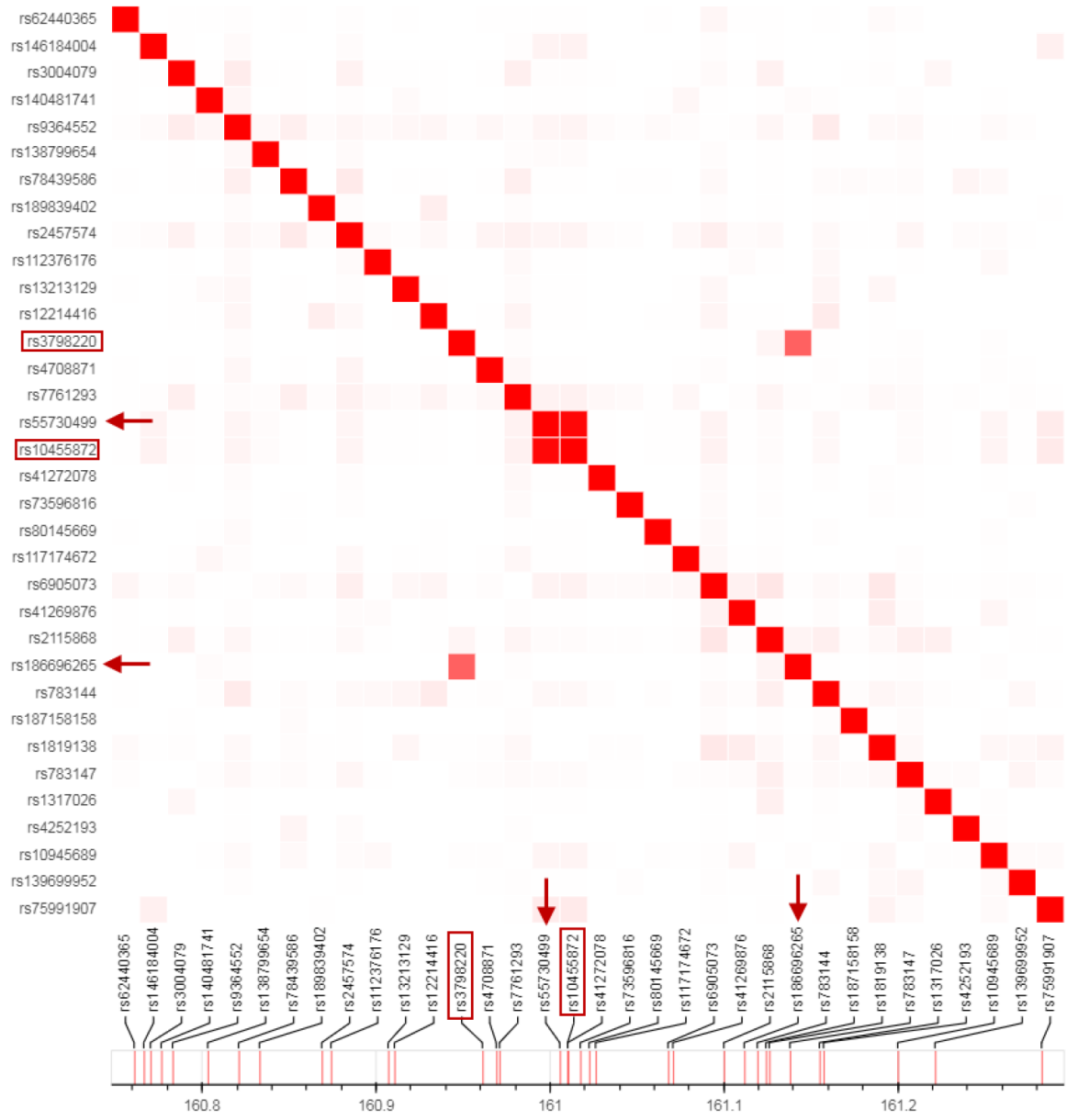
## Supplemental Figures and Figure Legends



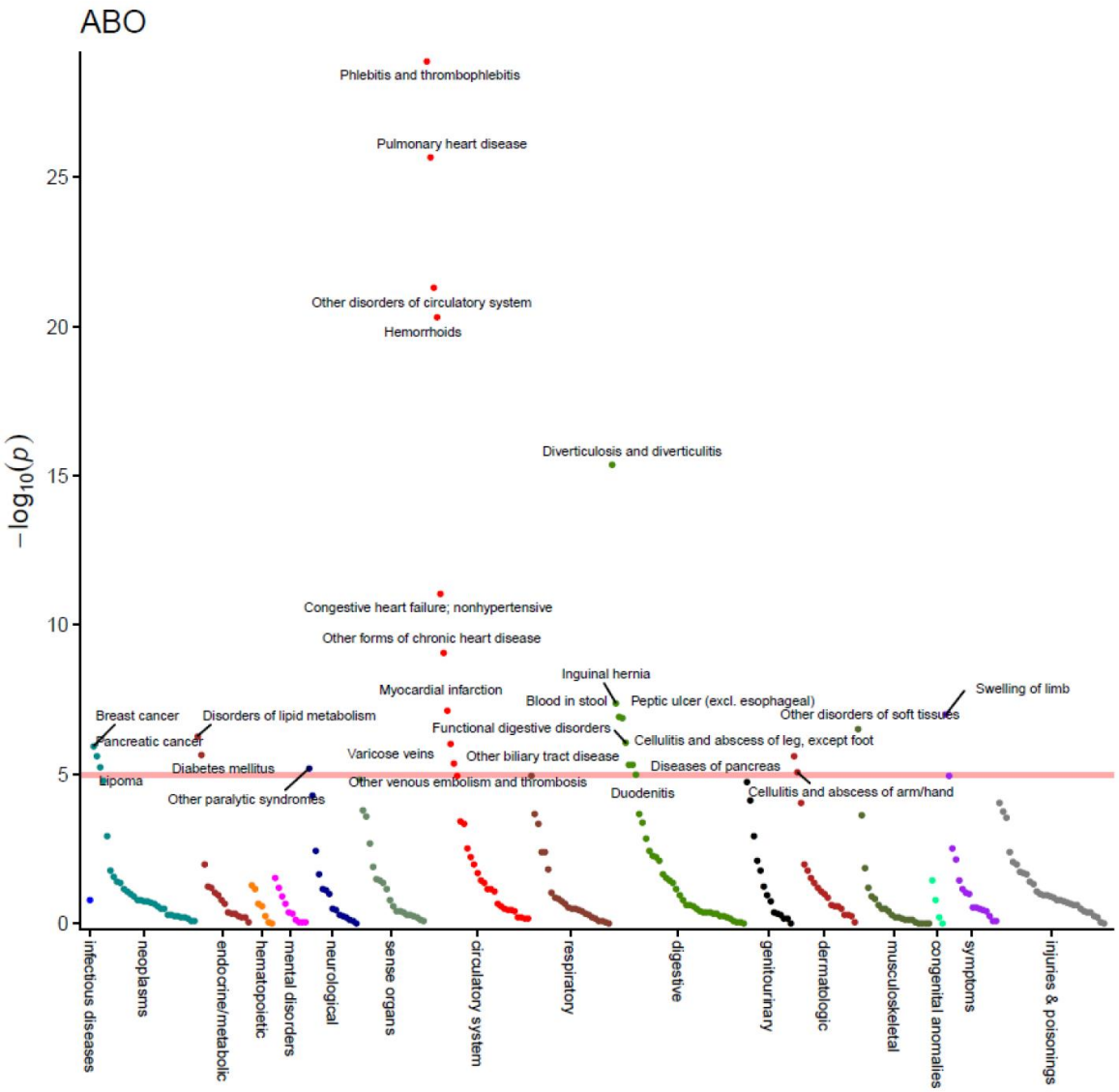

**Supplemental Figure 1**. **Study-specific MR results for significant biomarkers detected in multiple studies**. Other biomarkers were only measured in a single study and thus were not plotted. ABO and SCARA5 were measured in Sun *et al.* (2018), F11 in Suhre *et al.* (2017), and LPA in Sjaarda *et al.* (2018)[1–3].

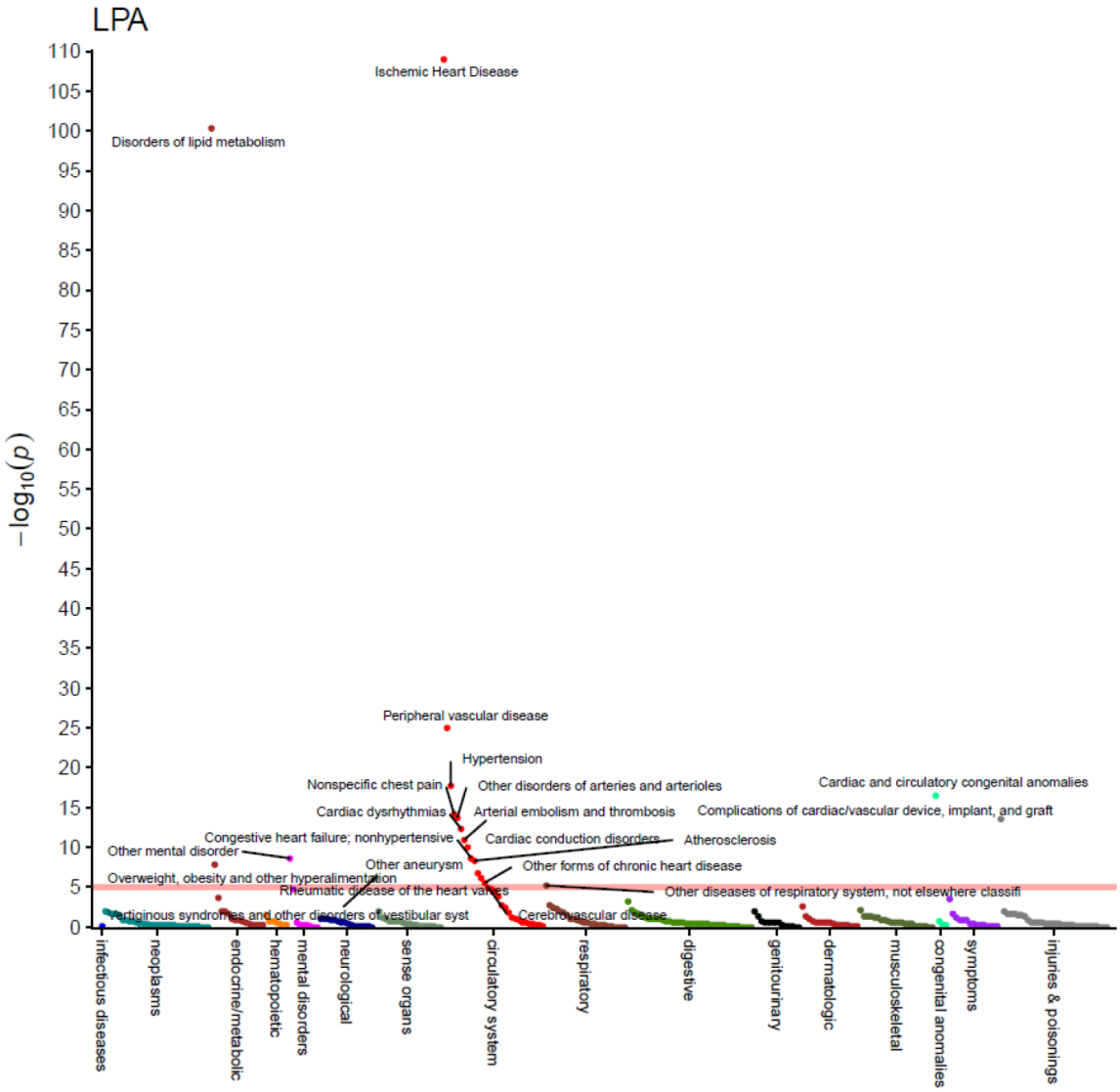


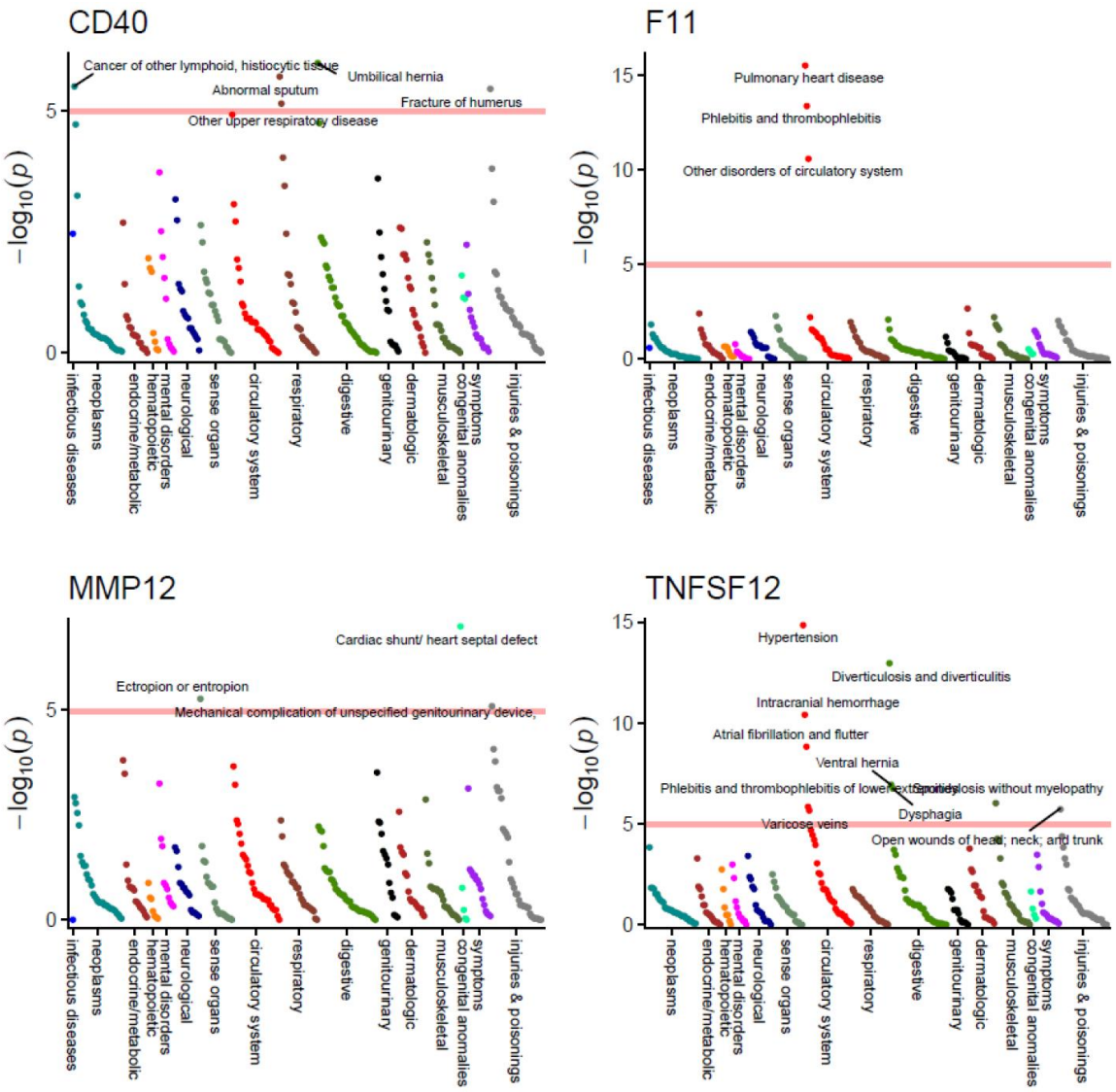

**Supplemental Figure 2. Linkage disequilibrium (LD) matrix between 6 SNPs selected as genetic instruments for F11 levels and 2 SNPs (rs4253417 and rs62350309) used as genetic instruments by Gill *et al.* (2018)[4]**. A red arrow indicates that a single SNP (rs2289252) in our study is strongly correlated (r2=0.97) with rs4253417 from Gill *et al.'s* (red box). After excluding this overlapping genetic signal, the association between F11 and CES remained statistically significant (OR=1.32; 95% CI, 1.16-1.50; $P=3.47 \times 10^{-5}$).

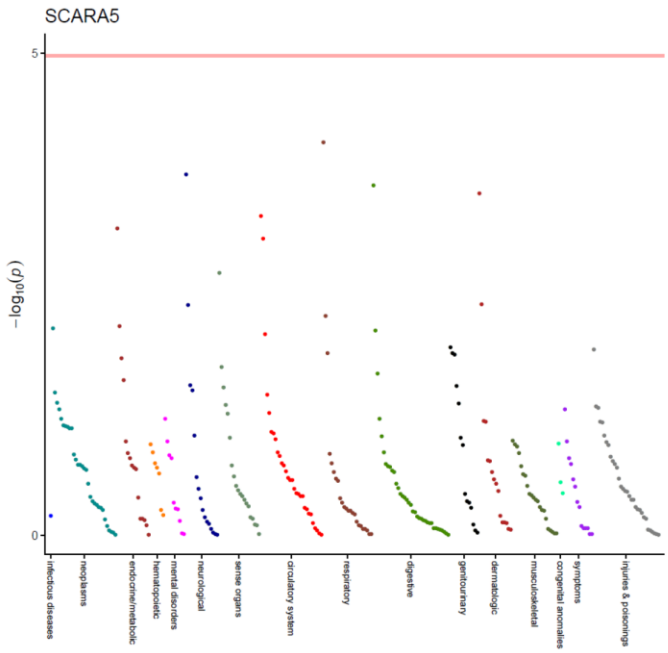



**Supplemental Figure 3. Linkage disequilibrium (LD) matrix between 31 SNPs selected as genetic instruments for LPA levels and two SNPs (rs3798220 and rs10455872) used in a genetic risk score by Helgadottir *et al.* (2012)**[5]. Red arrows indicate that two SNPs (rs55730499 and rs186696265) in our study are strongly correlated (r2>0.60) with SNPs in Helgadottir *et al.* (red box) After excluding overlapping genetic signals, the association between LPA and LAA remained statistically significant (OR=1.21; 95% CI, 1.11-1.32; P=1.93x10$^{-5}$).


ABO
Phlebitis and thrombophlebitis
Pulmonary heart disease
Other disorders of circulatory system
Hemorrhoids
Diverticulosis and diverticulitis
Congestive heart failure; nonhypertensive
Other forms of chronic heart disease
Myocardial infarction
Inguinal hernia
Blood in stool
Peptic ulcer (excl. esophageal)
Swelling of limb
Other disorders of soft tissues
Breast cancer
Disorders of lipid metabolism
Functional digestive disorders
Cellulitis and abscess of leg, except foot
Pancreatic cancer
Varicose veins
Other biliary tract disease
Diseases of pancreas
Diabetes mellitus
Other venous embolism and thrombosis
Cellulitis and abscess of arm/hand
Lipoma
Other paralytic syndromes
Duodenitis
$-\log_{10}(p)$
0
5
10
15
20
25
infectious diseases
neoplasms
endocrine/metabolic
hematopoietic
mental disorders
neurological
sense organs
circulatory system
respiratory
digestive
genitourinary
dermatologic
musculoskeletal
congenital anomalies
symptoms
injuries & poisonings


LPA
$-\log_{10}(p)$
Ischemic Heart Disease
Disorders of lipid metabolism
Peripheral vascular disease
Hypertension
Nonspecific chest pain
Other disorders of arteries and arterioles
Cardiac dysrhythmias
Arterial embolism and thrombosis
Cardiac conduction disorders
Atherosclerosis
Congestive heart failure; nonhypertensive
Other mental disorder
Other aneurysm
Other forms of chronic heart disease
Overweight, obesity and other hyperalimentation
Rheumatic disease of the heart valves
Other diseases of respiratory system, not elsewhere classifi
Vertiginous syndromes and other disorders of vestibular syst
Cerebrovascular disease
Cardiac and circulatory congenital anomalies
Complications of cardiac/vascular device, implant, and graft
infectious diseases
neoplasms
endocrine/metabolic
hematopoietic
mental disorders
neurological
sense organs
circulatory system
respiratory
digestive
genitourinary
dermatologic
musculoskeletal
congenital anomalies
symptoms
injuries & poisonings


CD40
Cancer of other lymphoid, histiocytic tissue
Umbilical hernia
Abnormal sputum
Fracture of humerus
Other upper respiratory disease
F11
Pulmonary heart disease
Phlebitis and thrombophlebitis
Other disorders of circulatory system
MMP12
Cardiac shunt/ heart septal defect
Ectropion or entropion
Mechanical complication of unspecified genitourinary device,
TNFSF12
Hypertension
Diverticulosis and diverticulitis
Intracranial hemorrhage
Atrial fibrillation and flutter
Ventral hernia
Phlebitis and thrombophlebitis of lower extremities
Spondylosis without myelopathy
Dysphagia
Varicose veins
Open wounds of head, neck, and trunk
$-\log_{10}(p)$
infectious diseases
neoplasms
endocrine/metabolic
hematopoietic
mental disorders
neurological
sense organs
circulatory system
respiratory
digestive
genitourinary
dermatologic
musculoskeletal
congenital anomalies
symptoms
injuries & poisonings




**Supplemental Figure 4. A series of manhattan plots illustrating significance of Phe-MR associations**. The horizontal pink line indicates the threshold for Bonferroni significance ($P<1.07x10^{-5}$), and labels are provided for disease traits surpassing this threshold. Disease traits are grouped and colour-coded by disease chapter and sorted by statistical significance. ABO = Histo-blood group ABO system transferase; CD40 = cluster of differentiation 40; F11 = coagulation factor XI; LPA = apolipoprotein(a); MMP12 = matrix-metalloproteinase-12; SCARA5 = scavenger receptor class A5; TNFSF12 = tumour necrosis factor-like weak inducer of apoptosis.

# Supplemental References


1. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P, Oliver-Williams C, Kamat MA, Prins BP, Wilcox SK, Zimmerman ES, Chi A, Bansal N, Spain SL, Wood AM, Morrell NW, Bradley JR, Janjic N, Roberts DJ, Ouwehand WH, Todd JA, Soranzo N, Suhre K, Paul DS, Fox CS, Plenge RM, Danesh J, Runz H, Butterworth AS. Genomic atlas of the human plasma proteome. *Nature*. 2018;558:73–79.
2. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, Sarwath H, Thareja G, Wahl A, Delisle RK, Gold L, Pezer M, Lauc G, Selim MAED, Mook-Kanamori DO, Al-Dous EK, Mohamoud YA, Malek J, Strauch K, Grallert H, Peters A, Kastenmüller G, Gieger C, Graumann J. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun*. 2017;8:14357.
3. Sjaarda J, Gerstein H, Chong M, Yusuf S, Meyre D, Anand SS, Hess S, Paré G. Blood CSF1 and CXCL12 as Causal Mediators of Coronary Artery Disease. *J Am Coll Cardiol*. 2018;72:300–310.
4. Gill D, Georgakis MK, Laffan M, Sabater-Lleal M, Malik R, Tzoulaki I, Veltkamp R, Dehghan A. Genetically Determined FXI (Factor XI) Levels and Risk of Stroke. *Stroke*. 2018;49:2761–2763.
5. Helgadottir A, Gretarsdottir S, Thorleifsson G, Holm H, Patel RS, Gudnason T, Jones GT, Van Rij AM, Eapen DJ, Baas AF, Tregouet DA, Morange PE, Emmerich J, Lindblad B, Gottster A, Kiemeny LA, Lindholt JS, Sakalihasan N, Ferrell RE, Carey DJ, Elmore JR, Tsao PS, Grarup N, Jørgensen T, Witte DR, Hansen T, Pedersen O, Pola R, Gaetani E, Magnadottir HB, Wijmenga C, Tromp G, Ronkainen A, Ruigrok YM, Blankensteijn JD, Mueller T, Wells PS, Corral J, Soria JM, Souto JC, Peden JF, Jalilzadeh S, Mayosi BM, Keavney B, Strawbridge RJ, Sabater-Lleal M, Gertow K, Baldassarre D, Nyyssnen K, Rauramaa R, Smit AJ, Mannarino E, Giral P, Tremoli E, De Faire U, Humphries SE, Hamsten A, Haraldsdottir V, Olafsson I, Magnusson MK, Samani NJ, Levey AI, Markus HS, Kostulas K, Dichgans M, Berger K, Kuhlenbumer G, Ringelstein EB, Stoll M, Seedorf U, Rothwell PM, Powell JT, Kuivaniemi H, Onundarson PT, Valdimarsson E, Matthiasson SE, Gudbjartsson DF, Thorgeirsson G, Quyyumi AA, Watkins H, Farrall M, Thorsteinsdottir U, Stefansson K. Apolipoprotein(a) genetic sequence variants associated with systemic atherosclerosis and coronary atherosclerotic burden but not with venous thromboembolism. *J Am Coll Cardiol*. 2012;60:722–729.
6. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, LeFaive J, VandeHaar P, Gagliano SA, Gifford A, Bastarache LA, Wei WQ, Denny JC, Lin M, Hveem K, Kang HM, Abecasis GR, Willer CJ, Lee S. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50:1335–1341.
7. Woo D, Falcone GJ, Devan WJ, Brown WM, Biffi A, Howard TD, Anderson CD, Brouwers HB, Valant V, Battey TWK, Radmanesh F, Raffeld MR, Baedorf-Kassis S, Deka R, Woo JG, Martin LJ, Haverbusch M, Moomaw CJ, Sun G, Broderick JP, Flaherty ML, Martini SR, Kleindorfer DO, Kissela B, Comeau ME, Jagiella JM, Schmidt H, Freudenberger P, Pichler A, Enzinger C, Hansen BM, Norrving B, Jimenez-Conde J, Giralt-Steinhauer E, Elosua R, Cuadrado-Godia E, Soriano C, Roquer J, Kraft P, Ayres AM, Schwab K, McCauley JL, Pera J, Urbanik A, Rost NS, Goldstein JN, Viswanathan A, Stögerer EM, Tirschwell DL, Selim M, Brown DL, Silliman SL, Worrall BB, Meschia JF, Kidwell CS, Montaner J, Fernandez-Cadenas I, Delgado P, Malik R, Dichgans M, Greenberg SM, Rothwell PM, Lindgren A, Slowik A, Schmidt R, Langefeld CD, Rosand J. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am J Hum Genet*. 2014;94:511–521.
8. Kjaergaard AD, Johansen JS, Bojesen SE, Nordestgaard BG. Elevated plasma YKL-40, lipids and lipoproteins, and ischemic vascular disease in the general population. *Stroke*. 2015;46:329–335.
9. Prins BP, Abbasi A, Wong A, Vaez A, Nolte I, Franceschini N, Stuart PE, Guterriez Achury J, Mistry V, Bradfield JP, Valdes AM, Bras J, Shatunov A, Lu C, Han B, Raychaudhuri S, Bevan S, Mayes MD, Tsoi LC, Evangelou E, Nair RP, Grant SFA, Polychronakos C, Radstake TRD, van Heel DA, Dunstan ML, Wood NW, Al-Chalabi A, Dehghan A, Hakonarson H, Markus HS, Elder JT, Knight J, Arking DE, Spector TD, Koeleman BPC, van Duijn CM, Martin J, Morris AP, Weersma RK, Wijmenga C, Munroe PB, Perry JRB, Pouget JG, Jamshidi Y, Snieder H, Alizadeh BZ. Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. *PLoS Med*. 2016;13:1–29.
10. Russi AE, Brown MA. Cystatin C and Cardiovascular Disease: A Mendelian Randomization Study. 2016;165:255–269.

# APPENDIX B:

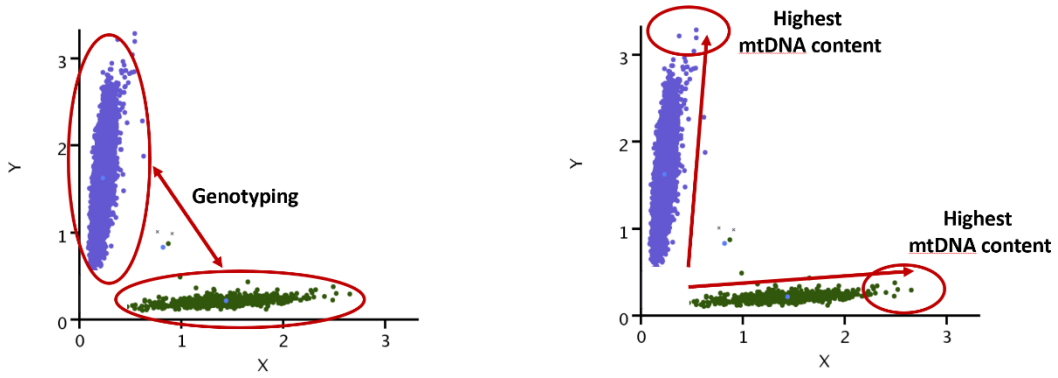
## Supplementary Data for Study 2

## Supplementary Results I:

## Development of the Automatic Mitochondrial Copy (AutoMitoC) Number Pipeline

### Background & Premise Underlying Array-based mtDNA-CN Estimation

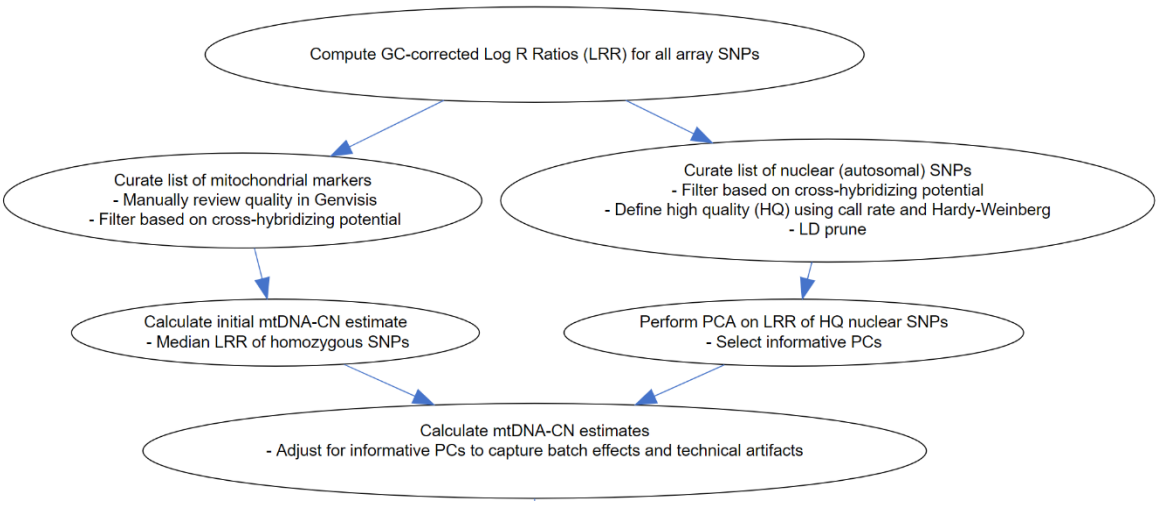
While SNP array data is intended for highly multiplexed determination of genotypes, the raw probe signal intensities used for genotypic inference can also be co-opted to derive estimates of mtDNA-CN. Determining genotypes for a sample at a given variant site relies on contrasting hybridization intensities of allele-specific oligonucleotide probes and then assigning membership to the most probable genotype cluster based on intensity properties (S1. Figure 1). Variation in intensities within each genotyping cluster can also be co-opted to deduce variations in copy number. A commonly used metric of probe intensity is the "log2ratio" (L2R), which denotes $\log_2$(observed intensity / expected intensity), where the expected intensity is defined as the median signal intensity for a probe conditional on each genotype cluster.




**S1. Figure 1.** Contrast in the intensities of mitochondrial probes X and Y discriminate genotypes. Intra-cluster variation in signal intensities may reflect mtDNA-CN. (Adapted from Lane *et al.*(Lane 2014))

**Existing Methodology: The MitoPipeline**

The "MitoPipeline" is a framework for estimating mtDNA-CN from array-based L2R (aka LRR) values developed by Lane *et al.* (2015) (Lane 2014; Zhang et al. 2017). An overview of the MitoPipeline is described in S1. Figure 2. To briefly summarize: (i) autosomal and mitochondrial L2R values are first corrected for GC waves; (ii) a high-quality set of mitochondrial and autosomal markers are selected largely based on visual inspection of genotype clusters and BLAST alignment for non-homologous sequences; (iii) principal component analysis (PCA) of at least 40,000 autosomal markers is conducted to capture batch effects; (iv) finally, mtDNA-CN is estimated based on median MT L2R value for each sample and then corrected for background noise through residualization of top autosomal PCs.

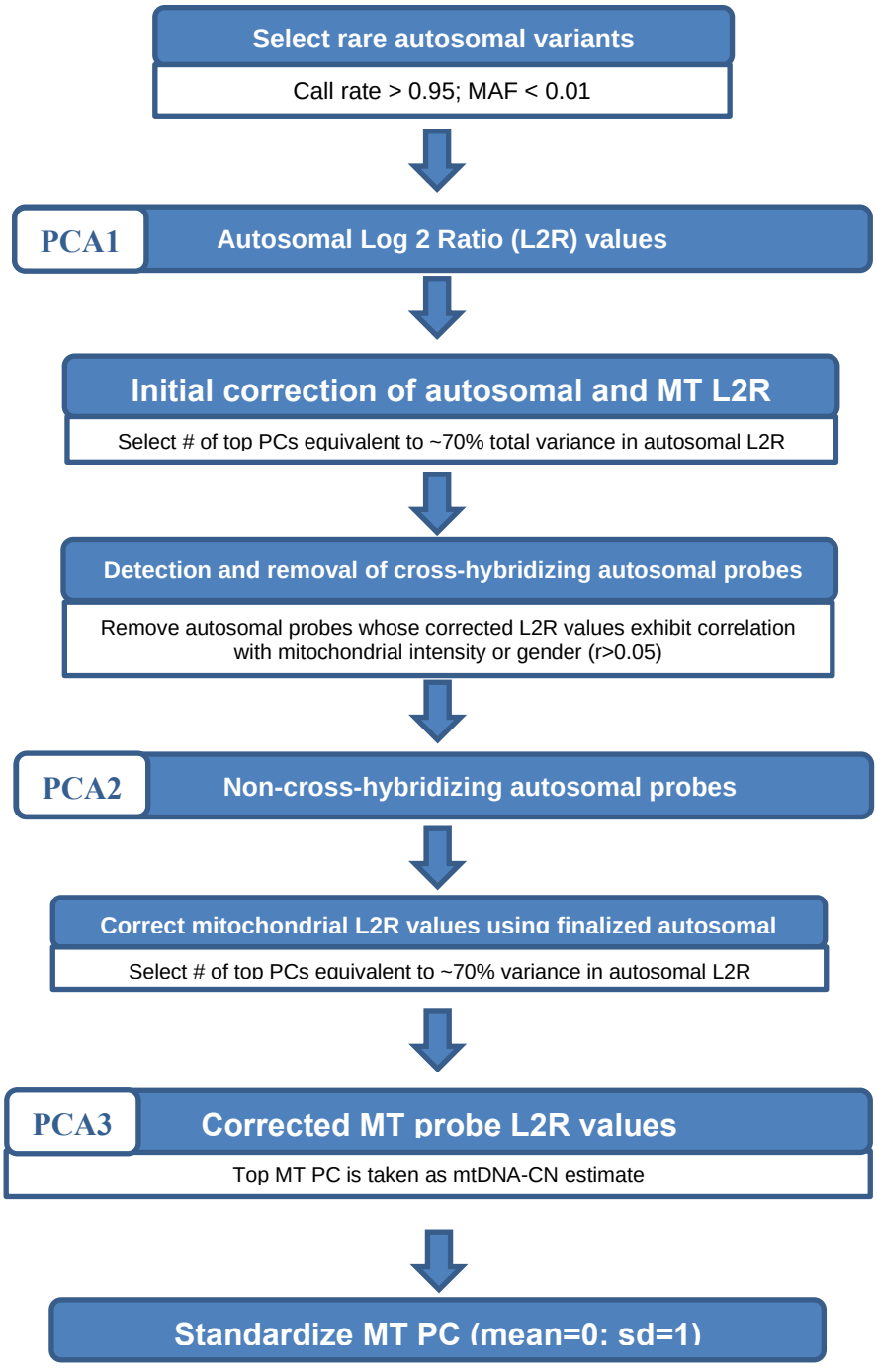



**S1. Figure 2.** Overview of the MitoPipeline (Source: http://genvisis.org/MitoPipeline/) (Lane 2014).

The MitoPipeline has proven to be effective in estimating mtDNA-CN (correlation coefficient R ~ 0.5 with direct qPCR estimates) as evidenced by multiple epidemiological studies employing this method (Ashar et al. 2017; Fazzini et al. 2019; Zhang et al. 2017). Firstly, visual inspection of MT probe intensity clusters is recommended to remove probes with poorly differentiated genotype clusters. However, this process is time-consuming (especially for biobank studies that are genotyped across thousands of batches); determination of probes with "good" vs "bad" genotype clustering is subjective; and guidance is only provided for adjudication of polymorphic but not monomorphic markers which may still be informative. Secondly, in consideration of nuclear and mitochondrial sequences with significant sequence similarity due to past and recurrent transposition of mitochondrial sequence into the nuclear genome, also known as "NUMTs" (Simone et al. 2011), the MitoPipeline recommends exclusion of MT probes with greater than 80% sequence similarity to the nuclear genome. While determining sequence homology of probes to the nuclear genome may have been feasible with older microarrays wherein probe sequences were often publicized, for many contemporary arrays, including the UKBiobank array, such information is not readily available. Thirdly, LD-pruning of common autosomal variants is required to ascertain a set of independent genetic variants, but implementation of this approach within ethnically diverse studies becomes more complex since genetic independence is ancestry-dependent. Under the MitoPipeline framework, each ethnicity warrants a unique set of common variants, which not only adds to computational burden but also creates an additional source of variability in terms of performance of the method between ethnicities.

**Proposed Methodology: AutoMitoC**

Therefore, we developed a new array-based mtDNA-CN estimation method, which we have dubbed the "AutoMitoC" pipeline, which incorporates three key amendments: (i) Autosomal signal normalization utilises globally rare variants in place of common variants which confers advantages in terms of both speed and portability to ethnically diverse studies, (ii) cross-hybridizing probes are identified by assessing evidence for cross-hybridization via association of signal intensities (rather than using genotype association and identification of homologous sequences through BLAST alignment), and (iii) the primary estimate of mitochondrial (MT) signal is ascertained using PCA as opposed to using the median signal intensity of MT probes. The rationale underlying these amendments are described in detail in the subsequent sections. An overview of the AutoMitoC pipeline is provided in S1. Figure 3.




**S1. Figure 3.** Overview of the AutoMitoC Pipeline.

**Development of AutoMitoC in the UKBiobank study**

To develop the AutoMitoC pipeline we used genetic datasets from the large UKBiobank prospective cohort study which includes approximately half of a million UK residents recruited from 2006 to 2010 in whom extensive genotypic and phenotypic investigations have been and continue to be performed (Sudlow et al. 2015). The size, breadth, and depth of such investigations makes this a rich resource for both methodological development and medical research. All UKBiobank data was accessed as part of application ID: 15255, "Identification of the shared biological and sociodemographic factors underlying cardiovascular disease and dementia risk". Two main genetic datasets from the UKBiobank were incorporated in the development of AutoMitoC. Firstly, CNV log2r (L2R) values derived from genetic arrays (i.e. normalized array probe intensities) for 488,264 samples were downloaded using the *ukbgene* utility, and their corresponding genotype calls (data field: 22418) were downloaded with *gfetch*. Secondly, exome alignment maps (EXOME FE CRAM files and indices; data fields 23163 & 23164) from the first tranche 49,989 samples released in March, 2019 were downloaded with the "ukbfetch" utility. L2R values were used to derive AutoMitoC mtDNA-CN estimates. For whole-exome sequencing data, Samtools *idxstats* was used to derive the number of sequence reads aligning to mitochondrial and autosomal genomes, from which an estimate of mtDNA-CN was derived according to the procedure by Longchamps *et al.* (2019) (Longchamps 2019); these complementary WES-based mtDNA-CN estimates served as a comparator to benchmark AutoMitoC performance.

Initial quality control of 488,264 samples and 784,256 directly genotyped variants was executed in PLINK following that of the Mitopipeline (i.e. sample call rate > 0.96; variant call rate > 0.98; HWE p-value > $1x10^{-5}$; PLINK mishap P-value > $1x10^{-4}$; genotype association with sex p-value > 0.00001; LD-pruning r2 < 0.30; MAF > 0.01) (Purcell et al. 2007). Variants within 1 Mb of immunoglobulin, T-cell receptor genes, and centromeric regions were removed. After this quality control procedure, 466,093 samples and 86,677 common variants remained. Next, genomic waves were corrected according to Diskin *et al.* (2008) using the PennCNV "genomic_wave.pl" script (https://github.com/WGLab/PennCNV/blob/master/genomic_wave.pl) (Diskin et al. 2008; Wang et al. 2007). Samples with high genomic waviness (L2R SD > 0.35) before and after GC-correction were removed resulting in 431,501 samples with array L2R values corresponding to 86,677 common autosomal variants. Lastly, we excluded samples representing blood cell count outliers as per Longchamps *et al.* (2019) which led to 395,781 participants (Longchamps 2019). Finally, we took the intersection of European samples with both suitable array and whole-exome sequencing data resulting in a final testing dataset of 34,436 European participants. To evaluate the possibility of replacing common autosomal variant signal normalization with rare variants, we also analyzed a set of 79,611 variants with a MAF≤0.01.

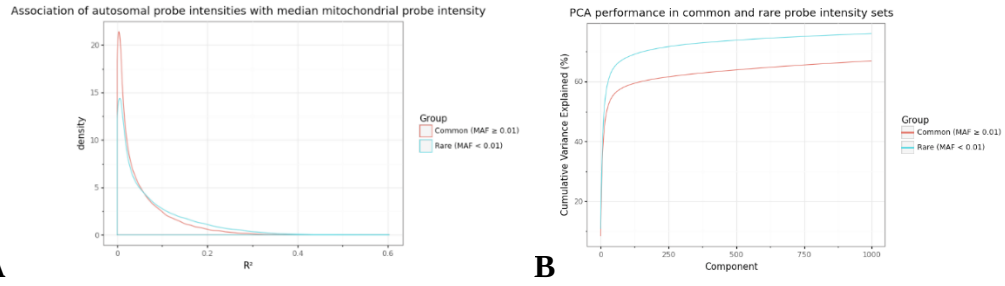
**Background correlation between autosomal & MT signal intensities**

Inference of relative mitochondrial DNA copy number (mtDNA-CN) from array data consists of determining the ratio of mitochondrial to autosomal probe signal intensities (or normalized L2R values) within each sample. Technical (e.g. batch and plate effects)

and latent sample factors confound raw signal intensities for reasons unrelated to DNA quantity. Such confounders induce strong cross-genome correlation, and therefore, a necessary first step is to remove this background noise from autosomal and mitochondrial probe intensities.

Indeed, even after correcting autosomal L2R values for genomic waves, we observed significant correlation between individual autosomal probe intensities and the median sample intensity across the 265 mitochondrial 2 (S1. Figure 4A). The extent of cross-genome intensity correlation varied based on minor allele frequency (MAF), with rare autosomal variants (MAF<0.01; M=79,611) showing the strongest correlation. We postulate that intensity properties for rare variants, which have a higher prevalence of homozygous genotypes, more strongly resemble those for mitochondrial genotypes, which are predominantly homoplasmic. On this basis, we only use rare autosomal variants to represent autosomal signal. While this approach contrasts with the Mitopipeline, which utilises common genetic variants to represent autosomal signal, restricting autosomal signal normalization to rare variants confers three major advantages while maintaining the same level of concordance with WES estimates ($R_{common}$=0.50 ; $R_{rare}$=0.49). First, this allows for further streamlining of the pipeline as this precludes the necessity for common variant filters, such as Hardy Weinberg equilibrium or LD-pruning. Second, we show that fewer principal components (PCs) are necessary to capture the same proportion of total variance in signal intensities with rare as opposed to common variants. Approximately 70% of the total variance in rare autosomal intensities was explained by 120 PCs, whereas the same proportion of variance in common autosomal intensities would necessitate more than 1000

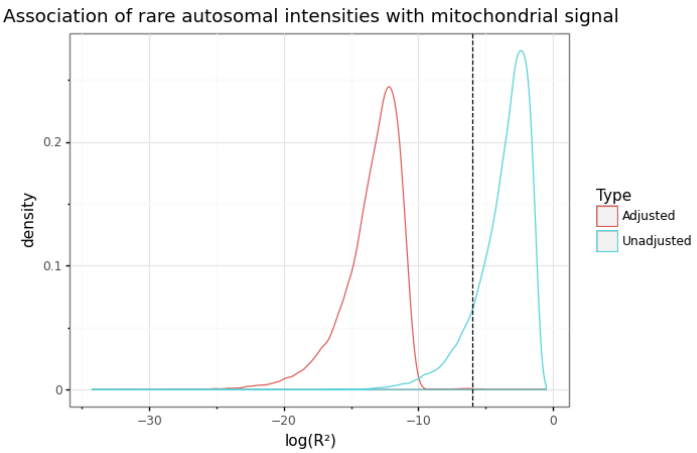
PCs (S1. Figure 4B). In the UKBiobank, PCs were derived via the eigendecomposition of the empirical covariance matrix conducted in Python 3.6, using NumPy and SciPy (Harris et al. 2020; Virtanen et al. 2020). Third, the set of autosomal markers used in deriving mtDNA-CN remains independent from the set of common autosomal variants analyzed in subsequent GWAS for mtDNA-CN. Effectively, this ensures that common autosomal variants evaluated for association with mtDNA-CN in downstream GWAS analyses are not directly incorporated into autosomal signal normalization, which could otherwise attenuate GWAS signals.




**A** **B**

**S1. Figure 4**. (A) Histogram illustrating the square of the Pearson correlation coefficient ($R^2$) of autosomal GC-corrected L2R values with median mitochondrial signal intensity stratified by MAF categories. (B) Cumulative variance explained by inclusion of top eigenvectors for sets of common (MAF>0.01; M=86,677) and rare (MAF≤0.01; M=79,611) autosomal probe sets.

**Empirical Detection of off-target probes**




**S1. Figure 5**. Distribution of log10 transformed coefficients of determination ($R^2$) from the association between autosomal probe intensities vs. median mitochondrial signal with (blue) or without (red) correction for background noise (i.e. 120 autosomal PCs). The dashed vertical line represents the threshold corresponding to "moderate" correlation ($|R|>0.05$ or $R^2>0.0025$), which is used to remove outlying probes that are associated with mitochondrial signal. Without correction for top PCs, the most autosomal probes exhibit some correlation with mitochondrial signal.

After adjustment for 120 autosomal PCs (approximating the elbow of the variance explained curve), there persisted a smaller subset of autosomal probes that were significantly correlated with median MT intensity (S1. Figure 5). We hypothesized that such probes either (i) cross-hybridize with the MT genome (i.e. lie within nuclear mitochondrial DNA (NUMT) regions) or (ii) corresponded to genetic loci involved in regulation of mtDNA-CN. As an illustration, S1. Table 1 conveys characteristics of the 10 most strongly correlated variants. Four of the top 10 probes were located within 1Mb of a NUMT region, and in all four cases, the sign of the correlation coefficient was negative

which might reflect interference of autosomal signal with increasing mtDNA-CN. An additional 3 probes corresponded to variants within genes that were implicated in mitochondrial disorders or regulation of mitochondrial processes (S1. Table 1).

**S1. Table 1**. Characteristics of top 10 autosomal probes whose adjusted intensities correlate with median mitochondrial signal.

| Autosomal Probe ID | Genomic Coordinates | Intensity R (Auto vs. Mito) | Comment |
|---|---|---|---|
| rs68130461 | 17:22024892 | -0.20 | HSA_NumtS_508_b2 |
| Affx-80229644 | 6:43484921 | -0.11 | HSA_NumtS_239_b1 (+25 Kb) |
| rs35201453 | 14:22783111 | -0.09 | NA |
| rs41267813 | 6:160998199 | -0.08 | HSA_NumtS_261_b1 (-742 Kb) |
| rs117507044 | 13:19958310 | -0.08 | NA |
| rs113200742 | 6:25272561 | -0.08 | HSA_NumtS_236_b1 (-324 Kb) |
| rs201397731 | 1:161168270 | 0.08 | *NDUFS2* Coding Variant (gene causes Mitochondrial Complex I Deficiency) |
| rs138167117 | 17:65739626 | 0.08 | NA |
| rs138656762 | 19:36330320 | 0.08 | *NPHS1* pathogenic variant (causes Finnish Nephrotic syndrome, characterized by mitochondrial dysfunction in kidneys) |
| rs117116233 | 19:8535980 | -0.08 | *HNRNPM* intronic variant (putative regulator of mitochondrial processes) |

We further explored whether there was evidence for cross-hybridization between autosomal SNPs and sex chromosomes by regressing adjusted autosomal probe intensities with reported male status. Generally, correlations between autosomal intensities and sex were stronger than those with MT intensities, suggesting that cross-hybridization of autosomal probes to sex chromosomes is more pronounced. For the top 10 sex-associated

probes, we performed BLASTn alignment against the human reference genome (GRCh38) using 30 bases surrounding each probe (S1. Table 2) (Altschul et al. 1990). All probes had at least one flanking sequence with near-perfect ($\geq$ 97%) sequence identity to a sex chromosome. For these 10 probes, the sign of the correlation between autosomal intensity and male status was perfectly consistent with homology to X or Y chromosomes, thus supporting the hypothesis that such probes cross-hybridized to sex chromosomes.

**S1. Table 2.** Characteristics of top autosomal probes associated with male sex status.

| Autosomal Probe ID | Genomic Coordinates | Intensity R (Auto vs. Male Sex) | Flank (30 BP) | Chromosome | Sequence Similarity (E-value) |
|---|---|---|---|---|---|
| Affx-89012246 | 7:141336763 | 0.67 | Right | Y | 97% ($9x10^{-6}$) |
| rs138167117 | 17:65739626 | -0.60 | Left | X | 100% ($3x10^{-8}$) |
| rs147585440 | 18:47310224 | -0.60 | Left | X | 100% ($3x10^{-8}$) |
| Affx-80264600 | 21:38555134 | -0.59 | Left | X | 100% ($3x10^{-8}$) |
| Affx-80229637 | 6:43470088 | -0.59 | Left | X | 100% ($3x10^{-8}$) |
| rs56275071 | 10:88822514 | -0.57 | Left | X | 97% ($9x10^{-6}$) |
| rs117507044 | 13:19958310 | 0.57 | Left | Y | 100% ($3x10^{-8}$) |
| Affx-89008518 | 7:140501303 | -0.57 | Left | X | 100% ($3x10^{-8}$) |
| Affx-80224967 | 4:84380892 | -0.57 | Right | X | 100% ($3x10^{-8}$) |
| Affx-89005068 | 7:140501288 | -0.55 | Left | X | 100% ($9x10^{-6}$) |

Inclusion of autosomal probes with evidence of off-target hybridization to the mitochondrial genome or sex chromosomes is problematic. In the former scenario,

misattribution of autosomal as mitochondrial signal may reduce the effectiveness of normalization. In the latter scenario, inadvertent adjustment for sex through retention of cross-hybridizing autosomal probes may occur if such probes explain substantial variance in autosomal probe intensities and this is particularly problematic given that mtDNA-CN has been robustly shown to differ between genders in epidemiological studies. In preliminary investigations where sex-associated probes were retained, we noticed that several top PCs that perfectly tagged gender. Hypothetically, had these PCs been retained and used for correction of mitochondrial signal, then the final mtDNA-CN estimate would have been inadvertently corrected for gender. Therefore, we removed autosomal probes exhibiting moderate correlation ($|R| > 0.05$) with sex (907; 1.14%) or median mitochondrial intensity (193; 0.24%) and then recalculated top autosomal PCs.

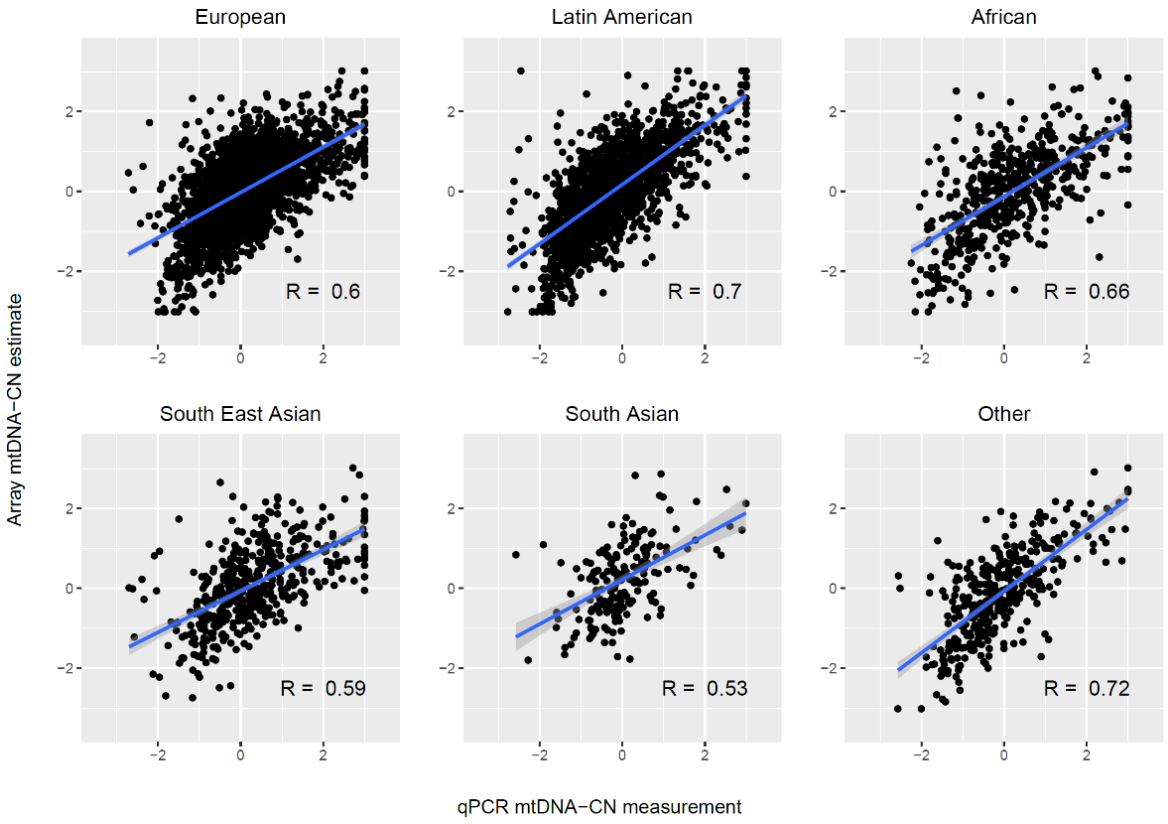
**PCA-based approach improves concordance with complementary estimates**

After correcting MT probes using the updated set of 120 autosomal L2R PCs, we adopted the Mitopipeline's approach for estimating mtDNA-CN and calculated the median of corrected MT L2R values to denote an individual's final mtDN A-CN estimate. Using this median-based approach, array mtDNA-CN estimates demonstrated significant correlation with WES mtDNA-CN estimates ($R=0.33$; $P < 2.23 \times 10^{-308}$). However, we found that performing PCA across all corrected MT L2R values and then extracting the top MT PC for each sample as the final mtDNA-CN estimate resulted in stronger correlation with WES ($R=0.49$; $P < 2.23 \times 10^{-308}$).

**Independent validation in an ethnically diverse sample**

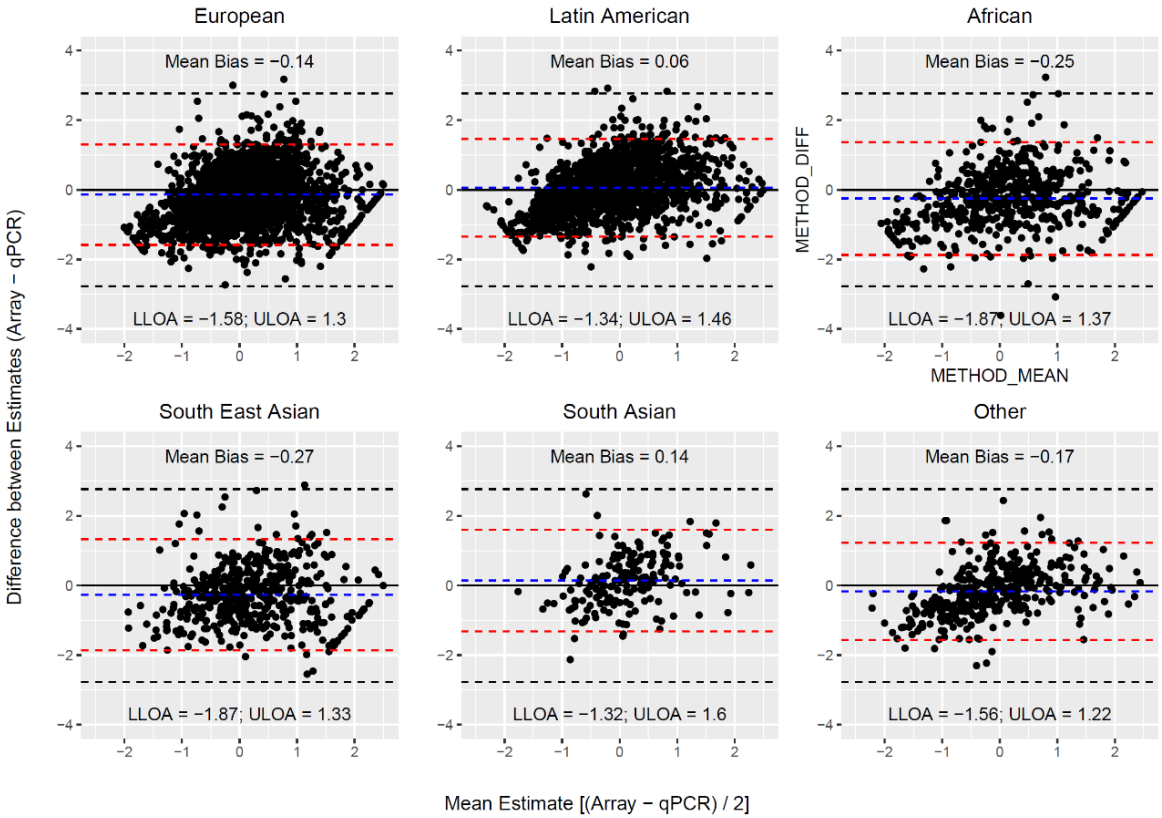
We additionally validated the AutoMitoC pipeline by deriving array-based mtDNA-CN estimates in the INTERSTROKE study and comparing these with parallel qPCR-based measurements, the current gold standard for measuring mtDNA-CN. INTERSTROKE is an international case-control study of stroke including 26,526 participants from 32 countries and 142 centers(O'Donnell et al. 2010). Blood samples have been collected for a subset of approximately 12,000 individuals, of which 9,311 have been successfully genotyped using the Axiom Precision Medicine Research Array (PMRA r3). A further subset of 5,791 samples with both suitable array genotypes have undergone qPCR measurement of mtDNA-CN using the plasmid-normalized protocol from Fazzini *et al.* (2019). Within INTERSTROKE, concordant findings to UKB-based analyses were observed favouring the PC-based ($r=0.64$; $P< 2.23x10^{-308}$) over the median-based approach ($R=0.60$; $P < 2.23x10^{-308}$). Furthermore, INTERSTROKE is ethnically diverse thus enabling an assessment of the robustness of AutoMitoC across genetic ancestries (S1. Figure 6). Correlations between array and qPCR mtDNA-CN estimates were comparable for individuals of European (N=2431), Latin American (N=1704), African (N=542), South East Asian (N=471), South Asian (N=186), and other ethnic groups (N=360; S1. Figure 6). Bland Altman plots also illustrate the extent of agreement between methods (S1. Figure 7). For every ethnicity, 95% limits of agreement intervals were smaller than expected by chance. Lastly, while all analyses hitherto followed the Mitopipeline condition of requiring > 40,000 autosomal variants for normalization, we observed comparable performance using even 1,000 random rare autosomal probes ($r2=0.60$; $P<5x10^{-300}$) for signal normalization

which reduced the runtime from several hours to less than 10 minutes for these 5,791 samples.




**S1. Figure 6.** Validation of AutoMitoC in an ethnically diverse cohort with qPCR-based estimates. Both qPCR and array-based mtDNA-CN estimates are presented as standardized units (mean=0; SD=1). The sample consisted of 2431 Europeans, 1704 Latin Americans, 542 Africans, 471 South East Asians, 186 South Asians, and 360 participants of other ancestry. Correlations between array and qPCR estimates were comparable for European (r=0.60; $P=2.7x10^{-238}$), Latin American (r=0.70; $P=3.9x10^{-251}$), African (R=0.66; $P=1.8x10^{-68}$), South East Asian (r=0.59; $P=6.2x10^{-46}$), South Asian (r=0.53; $P=4.2x10^{-15}$),

and other (r=0.72; P=5.4x10$^{-59}$) ethnic groups. The blue line indicates the linear trendline and the surrounding shaded region indicates the 95% confidence interval for the trendline.




**S1. Figure 7.** Bland Altman plots illustrating the extent of agreement between array and qPCR measurements. The black solid line indicates perfect agreement. The dashed blue line indicates the mean difference (or bias) between estimates. The horizontal red line corresponds to the 95% upper and lower limits of agreement (U/L LOA) for the observed data. The dashed black lines indicate the 95% U/L LOA that is expect under the null for two unrelated variables.

## Supplementary Methods

*** For all methods and results pertaining to the AutoMitoC pipeline, please see supplementary results I.

**The UKBiobank study**

The UKBiobank is a prospective cohort study including approximately 500,000 UK residents (ages 40-69 years) recruited from 2006-2010 in whom extensive genetic and phenotypic investigations have been and continue to be done(Sudlow et al. 2015). All UKBiobank data reported in this manuscript were accessed through the UKBiobank data showcase under application # 15525. All following analyses described in this supplementary material involve the use of genetic and/or phenotypic data from consenting UKBiobank participants.

**Genetic Analysis of Common Variants**

Data acquisition and quality control

UKBiobank samples were genotyped on either the UK Biobank Array (~450,000) or the UK BiLEVE array (~50,000) for approximately 800,000 variants (Bycroft et al. 2018). Further imputation was conducted by the UKBiobank study team using a combined reference panel of the UK10K and Haplotype Reference Consortium datasets. Imputed genotypes (version 3) for 488,264 UKBiobank participants were downloaded through the European Genome Archive (Category 100319). Samples were removed if they were flagged for any of the UKBiobank-provided quality control annotations (Resource 531; "ukb_sqc_v2.txt") for high ancestry-specific heterozygosity, high missingness, mismatching genetic ancestry, or sex chromosome aneuploidy ("het.missing.outliers","in.white.British.ancestry.subset",

"putative.sex.chromosome.aneuploidy"). Samples were also removed if their submitted gender did not match their genetic sex or if they had withdrawn consent at the time of analysis. Variant quality control consisted of removing variants that had low imputation quality (INFO score $\leq$ 0.30), were rare (MAF $\leq$ 0.005), or were in Hardy Weinberg Disequilibrium (HWE P $\leq 1x10^{-10}$). The HWE test was conducted within a subset of unrelated individuals for each ethnic strata, though all related individuals were retained for subsequent GWAS analysis. Lastly, in special consideration of mtDNA-CN as the GWAS phenotype, we also removed variants within "NUMTs", which refer to regions of the nuclear genome that exhibit homology to the mitochondrial genome due to past transposition of mitochondrial sequences. Accordingly, NUMTs represent a specific confounder of mtDNA-CN GWAS analyses which may lead to false positive associations. NUMT boundaries were obtained from the UCSC NumtS Sequence (numtSeq) track, which is based on the Reference Human NumtS curated by Simone *et al.* (2011) (Simone et al. 2011). All sample and variant quality control of imputed genotypes were executed using qctools and resultant bgen files were indexed with bgenix. Smaller ethnic groups with similar genetic ancestry were consolidated; individuals self-reporting as "African" or "Caribbean" were combined into a larger "African" stratum and individuals self-reporting as "Indian" or "Pakistani" were combined into a "South Asian" stratum. After quality control, 359689 British, 10598 Irish, 13189 Other White, 6172 South Asian, and 6133 African samples passing quality control also had suitable array-based mtDNA-CN estimates for subsequent GWAS analyses.

Genome-wide association study (GWAS)

GWAS were initially conducted in an ethnicity-stratified manner. The number of variants tested for association with mtDNA-CN varied for British (M=10,728,525), Irish (M=10,707,537), Other White (M=10,894,497), South Asian (M=11,350,981), and African (M=18,981,896) study participants, respectively. GWAS was performed using the REGENIE framework which consists of two steps (Mbatchou et al. 2020). In step 1, mtDNA-CN was predicted using a ridge regression model fit on a set of high-quality genotyped SNPs (MAF>0.01, MAC>100, genotype and sample missingness above 10%, and passing HWE ($p>10^{-15}$)) across the whole genome in blocks of 1000 SNPs. In step 2, the linear regression model was used to test the association of all SNPs adjusting for age, age$^2$, sex, chip type, 20 genetic principal components, and blood cell counts (white blood cell, platelet, and neutrophil counts), and conditional on the model from step 1.

Blood cell counts were determined for blood specimen collected at the initial assessment visit using Beckman Coulter LH750 analyzers (https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/haematology.pdf). Information on blood cell counts was retrieved from the UKBiobank data showcase. Individuals with missing values for any blood cell counts (~2.5%) were removed from any subsequent analysis involving blood cell counts. Quality control of blood counts was done following the same procedure as Longchamps *et al.* (2019) (Longchamps 2019). Except for platelet counts, all blood cell counts were log-transformed and samples exhibiting outlying values were removed (~4% samples). Lastly, values were standardized to have a mean of 0 and standard deviation of 1.

After ethnicity-specific GWAS were performed, results were combined through meta-analysis using METAL (Willer, Li, and Abecasis 2010). European (N=383,476) and trans-ethnic (N=395,718) GWAS meta-analyses were performed. We found that results from the trans-ethnic meta-analyses strongly resembled that of the European meta-analysis due to the high proportion of Europeans (97%). Accordingly, we report results from the European meta-analysis as the primary GWAS. To summarize statistical associations, Manhattan plots and quantile-quantile plots were generated by uploading summary statistics into the locus zoom web platform (https://my.locuszoom.org/) (Pruim et al. 2010). LD-score regression was performed to calculate the LD-score intercept by uploading GWAS results to the LDhub test center (http://ldsc.broadinstitute.org/) (Bulik-Sullivan et al. 2015). As per the instructions, all variants within the MHC region on chromosome 6 were removed prior to uploading. Annovar (version date 2020-06-07) was used to functionally annotate genome-wide significant loci based on their proximity (+/- 250kb) to genes (RefSeq), predicted effect on amino acid sequence, allele frequency in external datasets (1000Genomes), clinical pathogenicity (Clinvar), and in silico deleteriousness (CADD), and eQTL information (GTEx v8) (Abecasis et al. 2012; GTEX 2014; Landrum et al. 2015; Rentzsch et al. 2019).

*NUMT Sensitivity Analyses*

To assess whether genome-wide significant associations could be explained by cryptic mitochondrial pseudogenes (NUMTs), we performed sensitivity analyses as per Nandakumar *et al.* (2021)f. The MT genome was divided into thirds, and AutoMitoC estimates were rederived for each region (MT:1-6425; MT:6526-11947; MT:11948-16569)

using the corresponding MT variants belonging to these three consecutive regions. Association testing was performed for each region using REGENIE. For a given variant, if at least one region-based analysis yielded a non-significant association ($P<0.05$), we considered this as evidence of NUMT interference.

*Fine-mapping of GWAS signals*

We followed a similar protocol to Vuckovic *et al.* (2020) for fine-mapping mtDNA-CN loci (Vuckovic et al. 2020). All 9,602 genome-wide significant variants were consolidated into genomic blocks by grouping variants within 250kb of each other, yielding 72 distinct genomic blocks. LDstore was used to compute a pairwise LD correlation matrix for all variants within each block and across all samples included in the European GWAS meta-analysis (Benner et al. 2017). For each genomic block, FINEMAP was used to perform stepwise conditional regression, leading to 80 conditionally independent variants at genome-wide significance (Benner et al. 2016). The number of conditionally independent genetic signals per genomic block was used to inform the subsequent fine-mapping search parameters. Finally, the FINEMAP random stochastic search algorithm was applied to derive 95% credible sets constituting candidate causal variants that jointly contributed to 95% (or higher) of the posterior inclusion probabilities (Benner et al. 2016).

*Mitochondrial expression quantitative trait loci (mt-eQTL)*

Among our GWAS hits, we searched for mt-eQTLs using information from Ali *et al*. (2019), "Nuclear genetic regulation of the human mitochondrial transcriptome" (Ali et al. 2019). All variants in both Tables 1 and 2 were queried in the mtDNA-CN summary

statistics. When mt-eQTLs also had reported effect estimates, the consistency in direction-of-effects between mt-eQTL and mtDNA-CN associations was reported (S2. Table 3).

*Gene prioritization & pathway analyses*

The Data-driven Expression-prioritized Integration for Complex Traits (DEPICT) v.1.1 tool was used to map mtDNA-CN loci to genes based on shared co-regulation of gene expression (Pers et al. 2015). Genome-wide significant variants from the European GWAS meta-analysis were "clumped" into independent loci using PLINK "--clump-p1 5e-8 --clump-kb 500 --clump-r2 0.05" with LD correlation matrix derived from 1000Genomes Europeans (Purcell et al. 2007). DEPICT was subsequently run on independent SNPs using default settings. DEPICT identified 91 genes in total at a FDR of 0.05. Of the 91 genes, 4 non-coding genes were excluded from subsequent analyses for lack of a match in the GeneMANIA database(Warde-Farley et al. 2010). The excluded genes include a pseudogene (*PTMAP3*), an intronic transcript (*ALMS1-IT1*), and 2 long non-coding RNAs (*SNHG15*, *RP11-125K10.4*). The remaining 87 DEPICT-prioritized genes were uploaded to the GeneMANIA web platform (https://genemania.org/), which mines publicly available biological datasets to identify additional related genes based on functional associations (genetic interactions, pathways, co-expression, co-localization and protein domain homology). Based on the combined list of DEPICT and GeneMANIA identified genes, a network was formed in GeneMANIA maximizing the connectivity between all input genes using the default "Assigned based on query gene" setting to weight the network. Functional enrichment analysis was then performed to identify overrepresented Gene Ontology (GO) terms among all network genes (Gene and Consortium 2000). All network genes with at

least one GO annotation were compared to a background comprising all GeneMANIA genes with GO annotations.

*Mitochondrial annotation-based analyses*

To complement the previous analyses, we labelled prioritized genes with MitoCarta3 annotations and performed subsequent statistical enrichment analyses (Rath et al. 2021). MitoCarta3 is an exquisite database of mitochondrial protein annotations, which draws from mass spectrophotometry and GFP colocalization experiments of isolated mitochondria from 14 different tissues, as well as a plethora of other sources including literature review, to assign all human genes statuses indicating whether the corresponding proteins are expressed in the mitochondria or not. We tested whether prioritized genes were enriched for the mitochondrial proteome by using a binomial test in R. The number of "trials" was set to the total number of DEPICT and GeneMANIA-prioritized genes (107); the number of "successes" was set to the aforementioned gene subset that were labelled as mitochondrial proteins by MitoCarta3 (27); finally, the expected probability was set to the number of nuclear-encoded MitoCarta3 genes divided by the total number of genes (1120/18922). Furthermore, a t-test was used to compare mean PGC-1A induced fold change for the 27 genes as compared to the mean PGC-1A induced fold change for all 1120 nuclear MitoCarta3-annotated genes. MitoCarta3 genes with missing values were excluded from this analysis. Lastly, the 27 genes were labelled based on MitoCarta3 "MitoPathways". Only the top level pathway (i.e. parent node) was ascribed to each gene within the main text though detailed pathway annotations are available S2. Table 5.

**Genetic Analysis of Rare Variants**

*Data acquisition and quality control*

Population-level whole-exome sequencing (WES) variant genotypes (UKB data field: 23155) for 200,643 UKBiobank participants corresponding to 17,975,236 variants were downloaded using the gfetch utility. These data represent the second tranche of WES data released by the UKBiobank and differs from the first tranche (~50K samples) which was used for the development of the AutoMitoC pipeline. Quality control of WES data was conducted as follows. First, 11 samples who withdrew consent by the time of analysis were removed. Second, 83,700 monomorphic variants were removed. Third, 369,215 variants with non-missing genotypes present in less than 90% of samples were removed. Fourth, 2 samples with call rates less than 99% were removed. Fifth, 18 samples exhibiting discordance between genetic and reported sex were removed. Sixth, through visual inspection of scatterplots of the first two genetic principal components, 3 outlying samples whose locations strongly departed from their putative ethnicity cluster were removed. Seventh, 35,317 variants deviating from Hardy Weinberg Equilibrium were removed. Eighth, 12,765 samples belonging to smaller ethnic groups with less than 5000 samples (South Asian=3395; African=3168; Other=6,202) were removed. Ninth, we selected for a maximal number of unrelated samples and excluded 14,156 samples exhibiting third degree or closer relatedness. Finally, 12,394,404 non-coding variants were removed, and 5,176,300 protein-altering variants (stopgain, stoploss, startloss, splicing, missense, frameshift and in-frame indels) were retained in 173,688 samples.

*Exome-wide association testing to identify rare mtDNA-CN loci*

Of the 173,688 individuals passing quality control, 147,740 had non-missing mtDNA-CN estimates. Further variant inclusion criteria were implemented: variants that were rare (MAF $\leq$ 0.001), non-synonymous, and predicted to be clinically deleterious by Mendelian Clinically Applicable Pathogenicity (M-CAP) v.1.4 scores (or were highly disruptive variant types including frameshift indel, stopgain, stoploss, or splicing) were retained (Jagadeesh et al. 2016). Herein, such variants are referred to as "rare variants" for simplicity. For each gene, rare allele counts were added per sample. 18,890 genes with a total minor allele count of least 10 were subsequently analyzed (exome-wide significance $P < 0.05/18890 = 2.65 \times 10^{-6}$). Linear regression was conducted using mtDNA-CN as the dependent variable and the rare alleles counts per gene as the independent variable. The same set of covariates used in the primary GWAS were also employed in this analysis.

*Phenome-wide association testing for rare SAMHD1 mutation carrier status*

To identify disease phenotypes associated with carrying a rare *SAMHD1* mutation, we maximized sample size for phenome-wide association testing by analyzing the larger set of 173,688 WES samples (with or without suitable mtDNA-CN estimates). Disease outcomes were defined using the previously published "PheCode" classification scheme to aggregate ICD-10 codes from hospital episodes (field ID 41270), death registry (field ID 40001 and 40002), and cancer registry (field ID 40006) records(Denny et al. 2013; Wu et al. 2019). Further manual review was performed to exclude cases of sex-specific outcomes that may be erroneously attributed to the opposite genetic sex. Logistic regression was applied to test the association of *SAMHD1* mutation carrier status versus 771 PheCodes (phenome-wide significance $P < 0.05/771 = 6.49 \times 10^{-5}$) with a minimal case sample size of

300 (Wei et al. 2017). The same set of covariates used in the primary GWAS were also employed in this analysis.

**Mendelian Randomization Analysis**

*Disease Outcomes*

To assess evidence for a causal role of mtDNA-CN on mitochondrial disorder-related traits, we first defined a list of testable disease outcomes related to mitochondrial disorders. 36 clinical manifestations from a review paper by Gorman *et al.* (2016) were cross-referenced to GWAS traits analyzed by the FinnGen consortium (Feng et al. 2020; Gorman et al. 2016). The FinnGen consortium is a collaborative research entity aggregating genomic data from 9 Finnish biobanks with phenotypic data from electronic health records (https://finngen.gitbook.io/documentation/data-description#summary-association-statistics). FinnGen GWAS (v4) have been performed for 176,899 participants and 2,444 disease endpoints using the SAIGE method which entails a logistic mixed model with saddle point approximation to account for imbalanced case-control ratios(Zhou et al. 2018). Of these 2,444 disease endpoints, 10 traits corresponded to one of the 36 clinical manifestations of mitochondrial disease and had a case prevalence greater than 1% in FinnGen including type 2 diabetes (N=23,364), mood disorder (N=20,288), sensorineural hearing loss (N=12,550), cerebrovascular disease (N=10,367), migraine (N=6,687), dementia (5,675), epilepsy (N=4,558), paralytic ileus and intestinal obstruction (N=2,999), and cardiomyopathy (N=2,342). Genome-wide summary statistics were downloaded for these 10 traits, from which effect estimates and standard errors were used in subsequent

Mendelian randomization analyses to define the effect of selected genetic instruments on disease risk.

*Genetic Instrument Selection*

First, genome-wide significant variants from the European GWAS meta-analysis of mtDNA-CN were chosen (N=383476). Second, we matched these variants to the FinnGen v4 GWAS datasets (Feng et al. 2020). Third, to enrich for variants that directly act through mitochondrial processes, we only retained those within 100kb of genes encoding for proteins that localize to the mitochondria based on MitoCarta3 annotations (Rath et al. 2021). Fourth, we performed LD-pruning in PLINK with 1000Genomes Europeans as the reference panel to ascertain an independent set of genetic variants (LD r2 > 0.01), resulting in 34 variants (Abecasis et al. 2012; Purcell et al. 2007). Lastly, to mitigate potential for horizontal pleiotropy, we further removed variants with strong evidence of acting through alternative pathways by performing a phenome-wide search across published GWAS with Phenoscanner V2 (Kamat et al. 2019). Variants strongly associated with other phenotypes ($P<5x10^{-20}$) were removed unless the variant was a coding mutation located within gene encoding for the mitochondrial proteome (MitoCarta3) or had an established mitochondrial role based on manual literature review (Rath et al. 2021). Seven genetic variants were removed based on these criteria including rs8067252 (*ADAP2*), rs56069439 (*ANKLE1*), rs2844509 (*ATP6V1G2-DDX39B*), rs73004962 (*PBX4*), rs7412 (*APOE*), rs385893 (*AK3, RCL1*), and rs1613662 (*GP6*) (S2. Table 8).

*Mendelian Randomization & Sensitivity Analyses*

Two sample Mendelian Randomization analyses were performed using the "TwoSampleMR" and "MRPRESSO" R packages (Hemani et al. 2018; Verbanck et al. 2018). Effect estimates and standard errors corresponding to the 27 genetic variants on mtDNA-CN (exposure) and mitochondrial disease phenotypes (outcome) were derived from the European GWAS meta-analysis and FinnGen v4 GWAS summary statistics, respectively (S2. Table 9). Three MR methodologies were employed including Inverse Variance Weighted (primary method), Weighted Median, and MR-EGGER methods. MR-PRESSO was used to detect global heterogeneity and P-values were derived based on 1000 simulations. If significant global heterogeneity was detected ($P<0.05$), a local outlier test was conducted to detect outlying SNPs. After removal of outlying SNPs, MR analyses were repeated. In the absence of heterogeneity (Egger-intercept $P \geq 0.05$; MR-PRESSO global heterogeneity $P \geq 0.05$), we reported the inverse-variance weighted result. In the presence of balanced pleiotropy (MR-PRESSO global heterogeneity $P < 0.05$) and absence of directional pleiotropy (Egger-intercept $P \geq 0.05$), we reported the weighted median result. In the presence of directional pleiotropy (Egger-intercept $P < 0.05$), we reported the MR-EGGER result. We also performed the Steiger directionality test to ensure that a greater proportion of variance in mtDNA-CN was explained than risk of the outcome. Finally, to replicate the two-sample MR finding using an independent outcome dataset without UKBiobank participants, we repeated two-sample MR analyses using the International Genomics of Alzheimer's Disease Consortium (2013) GWAS meta-analysis including 17,008 cases and 37,154 controls (Lambert et al. 2013).

## Supplementary Results II

# Supplementary Tables

**S2. Table 1**. Annotated genome-wide significant mtDNA-CN loci

| Variant Coordinates | | | | | | GWAS Meta-analysis Results | | | | NUMT Sensitivity Analysis Results | | | | | Gene Annotations | | | Allele Frequency |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| locus # | rsid | chr | pos | ref | alt | beta | se | pvalue | het_p | p_NUMT474 | p_NUMT228 | p_MT_1ST_THIRD | p_MT_2ND_THIRD | p_MT_3RD_THIRD | Gene.refGene | mtDNA Depletion Gene | Human MitoCarta3 Gene | 1000g2015aug_eur |
| 1 | rs2977608 | 1 | 768253 | A | C | 0.0236 | 0.0025 | 2.58E-21 | 0.2222 | 5.66E-12 | 4.75E-40 | 3.07E-27 | 8.92E-41 | 6.59E-01 | LINC01128 | No | No | 0.7783 |
| 2 | rs1569419 | 1 | 2996602 | T | C | 0.0189 | 0.0025 | 6.87E-14 | 0.5991 | 4.96E-14 | 3.10E-14 | 3.08E-13 | 7.92E-10 | 5.04E-11 | PRDM16 | No | No | 0.7555 |
| 3 | rs3766744 | 1 | 12043717 | G | A | -0.0179 | 0.0021 | 2.91E-17 | 0.8234 | 1.17E-17 | 2.48E-16 | 1.21E-15 | 3.32E-11 | 2.24E-15 | MFN2 | No | Yes | 0.4722 |
| 4 | rs139898146 | 1 | 25600747 | C | A | 0.0468 | 0.0021 | 1.96E-106 | 0.1535 | 9.15E-111 | 4.30E-86 | 2.96E-88 | 4.47E-46 | 7.80E-118 | RHD;RSRP1 | No | No | NA |
| 5 | rs2274319 | 1 | 156450873 | T | C | -0.0139 | 0.0022 | 4.09E-10 | 0.8799 | 1.08E-09 | 2.94E-10 | 2.24E-10 | 7.96E-05 | 2.08E-08 | MEF2D | No | No | 0.675 |
| 6 | rs2038480 | 1 | 171939641 | A | T | 0.0162 | 0.0027 | 1.39E-09 | 0.1239 | 7.71E-09 | 2.33E-09 | 2.09E-08 | 6.77E-07 | 1.50E-05 | DNM3 | No | No | 0.832 |
| 6 | rs143989240 | 1 | 172381181 | T | TTG | 0.0127 | 0.0023 | 2.61E-08 | 0.1015 | 4.16E-08 | 2.77E-08 | 9.49E-09 | 4.82E-05 | 5.49E-06 | DNM3 | No | No | 0.3101 |
| 7 | rs10749636 | 1 | 248020448 | G | A | 0.0155 | 0.0025 | 5.43E-10 | 0.6798 | 4.77E-10 | 5.59E-10 | 3.51E-09 | 2.66E-06 | 1.55E-09 | TRIM58 | No | No | 0.7694 |
| 8 | rs116640044 | 2 | 68694364 | C | T | -0.0691 | 0.0124 | 2.28E-08 | 0.4687 | 3.50E-08 | 6.56E-09 | 3.82E-09 | 4.02E-04 | 1.63E-06 | FBXO48 | No | No | 0.0089 |
| 9 | rs1653261 | 2 | 74158044 | T | C | -0.0246 | 0.0022 | 3.56E-28 | 0.4785 | 3.41E-28 | 1.83E-27 | 4.51E-27 | 2.42E-16 | 1.69E-24 | DGUOK | Yes | Yes | 0.6332 |
| 9 | rs62641680 | 2 | 74166053 | G | A | -0.0903 | 0.0063 | 4.63E-47 | 0.3075 | 1.44E-46 | 1.10E-46 | 8.38E-45 | 1.02E-27 | 3.27E-37 | DGUOK | Yes | Yes | 0.0249 |
| 9 | rs74874677 | 2 | 74177777 | A | G | -0.0819 | 0.0071 | 3.62E-31 | 0.6767 | 4.02E-31 | 2.91E-30 | 3.47E-30 | 1.10E-20 | 9.38E-24 | DGUOK | Yes | Yes | 0.0298 |
| 10 | rs12052715 | 2 | 160677375 | C | G | -0.0133 | 0.0024 | 1.72E-08 | 0.0227 | 2.76E-08 | 1.62E-08 | 8.67E-08 | 6.18E-06 | 1.66E-07 | LY75;LY75-CD302 | No | No | 0.6938 |
| 11 | rs78909033 | 2 | 241510903 | G | A | -0.021 | 0.0031 | 1.00E-11 | 0.4258 | 8.55E-12 | 1.20E-11 | 3.05E-12 | 1.83E-06 | 4.01E-08 | RNPEPL1 | No | No | 0.1233 |
| 12 | rs13084580 | 3 | 39188182 | C | T | 0.0244 | 0.0033 | 2.19E-13 | 0.9682 | 3.32E-13 | 5.38E-14 | 3.62E-12 | 9.07E-11 | 8.22E-10 | CSRNP1 | No | No | 0.1014 |
| 13 | rs6792510 | 3 | 48723302 | G | C | -0.0126 | 0.0022 | 1.44E-08 | 0.4732 | 1.00E-08 | 2.43E-08 | 1.33E-08 | 2.92E-05 | 7.43E-07 | NCKIPSD | No | No | 0.6471 |
| 14 | rs1354034 | 3 | 56849749 | T | C | -0.0268 | 0.0022 | 2.05E-35 | 0.6397 | 1.20E-35 | 7.46E-37 | 3.32E-37 | 3.40E-18 | 3.97E-24 | ARHGEF3 | No | No | 0.5994 |
| 15 | rs34778241 | 3 | 71771215 | T | TG | 0.014 | 0.0023 | 2.02E-09 | 0.9757 | 3.02E-09 | 2.65E-09 | 7.07E-09 | 3.93E-06 | 1.44E-07 | EIF4E3 | No | No | 0.6183 |
| 16 | rs13088724 | 3 | 179152841 | G | A | 0.0175 | 0.0026 | 7.70E-12 | 0.0437 | 1.28E-11 | 1.08E-10 | 8.75E-10 | 7.40E-09 | 9.97E-11 | GNB4 | No | No | 0.2584 |
| 17 | rs4698839 | 4 | 102801941 | C | T | 0.0124 | 0.0022 | 9.47E-09 | 0.5554 | 5.91E-09 | 3.93E-08 | 5.52E-08 | 2.87E-05 | 9.50E-09 | BANK1 | No | No | 0.3807 |
| 18 | rs7705526 | 5 | 1285974 | C | A | 0.0178 | 0.0023 | 6.46E-15 | 0.8712 | 1.14E-14 | 2.57E-15 | 5.61E-15 | 3.16E-11 | 7.48E-10 | TERT | No | No | 0.337 |
| 19 | rs36009521 | 5 | 72046100 | T | TA | 0.0181 | 0.0031 | 6.27E-09 | 0.2324 | 7.47E-09 | 3.14E-09 | 1.75E-08 | 1.60E-07 | 8.96E-07 | LINC02056;TNPO1 | No | No | 0.1382 |
| 20 | rs114694170 | 5 | 88180196 | T | C | 0.0331 | 0.0045 | 3.04E-13 | 0.8778 | 5.82E-14 | 2.20E-12 | 1.35E-11 | 2.62E-10 | 5.19E-11 | MEF2C-AS1 | No | No | 0.0537 |
| 21 | rs212930 | 6 | 25535636 | A | G | 0.0149 | 0.0025 | 3.18E-09 | 0.6535 | 4.16E-09 | 3.94E-09 | 3.42E-09 | 8.21E-06 | 4.32E-06 | CARMIL1 | No | No | 0.7763 |
| 22 | rs2844509 | 6 | 31510924 | A | G | 0.014 | 0.0025 | 2.61E-08 | 0.0344 | 2.01E-08 | 6.12E-09 | 1.42E-08 | 7.45E-06 | 1.55E-05 | ATP6V1G2-DDX39B | No | No | 0.2704 |
| 23 | rs35372932 | 6 | 32564985 | C | T | -0.0251 | 0.0043 | 4.12E-09 | 0.6261 | 1.54E-09 | 4.91E-08 | 1.13E-07 | 1.74E-05 | 1.16E-09 | HLA-DRB1;HLA-DQA1 | No | No | 0.0785 |
| 24 | rs5745582 | 6 | 33546498 | C | T | 0.021 | 0.0028 | 3.46E-14 | 0.7901 | 3.19E-14 | 2.13E-13 | 8.26E-13 | 7.53E-11 | 7.59E-11 | BAK1 | No | Yes | 0.2247 |
| 25 | rs4895441 | 6 | 135426573 | A | G | 0.0177 | 0.0024 | 8.31E-14 | 0.1309 | 1.68E-13 | 1.03E-13 | 1.77E-12 | 1.14E-10 | 6.59E-10 | HBS1L;MYB | No | No | 0.2734 |
| 26 | rs2304693 | 7 | 45148667 | G | A | 0.0181 | 0.0028 | 5.11E-11 | 0.2472 | 5.95E-11 | 4.77E-11 | 6.12E-11 | 2.21E-06 | 9.26E-09 | TBRG4 | No | Yes | 0.161 |
| 27 | rs11764390 | 7 | 80216205 | G | A | -0.0122 | 0.0022 | 1.55E-08 | 0.1394 | 1.55E-08 | 4.15E-08 | 1.31E-07 | 3.54E-06 | 2.42E-06 | GNAT3;CD36 | No | No | 0.4543 |
| 28 | rs445 | 7 | 92408370 | C | T | 0.0207 | 0.0036 | 8.85E-09 | 0.5592 | 5.10E-09 | 2.97E-08 | 3.61E-08 | 1.21E-05 | 5.40E-07 | CDK6 | No | No | 0.1143 |
| 29 | rs342293 | 7 | 106372219 | C | G | 0.0283 | 0.0021 | 1.05E-40 | 0.9779 | 1.04E-40 | 9.24E-41 | 2.61E-38 | 1.30E-24 | 5.82E-30 | CCDC71L;PIK3CG | No | No | 0.4374 |
| 30 | rs6972244 | 7 | 114716286 | G | A | 0.0355 | 0.0034 | 2.77E-25 | 0.9694 | 2.64E-22 | 6.95E-44 | 1.04E-67 | 3.35E-05 | 4.16E-01 | MDFIC;LINC01393 | No | No | 0.0815 |
| 31 | rs6959832 | 7 | 135289854 | G | A | -0.021 | 0.0021 | 2.10E-23 | 0.2241 | 4.58E-23 | 1.56E-24 | 4.09E-23 | 1.03E-15 | 5.64E-16 | NUP205 | No | No | 0.5139 |
| 32 | rs4284061 | 8 | 6878257 | T | A | -0.0186 | 0.0022 | 2.97E-17 | 0.5294 | 5.16E-17 | 7.60E-17 | 3.66E-15 | 4.03E-15 | 3.45E-12 | DEFA3;DEFA11P | No | No | 0.4573 |
| 33 | rs4841132 | 8 | 9183596 | A | G | -0.0212 | 0.0037 | 7.46E-09 | 0.8245 | 7.08E-09 | 1.36E-08 | 5.09E-10 | 1.98E-05 | 4.02E-05 | LOC157273 | No | No | 0.9264 |
| 34 | rs2322718 | 8 | 27257787 | T | G | 0.0136 | 0.0021 | 1.53E-10 | 0.7564 | 1.35E-10 | 1.35E-09 | 2.06E-09 | 6.87E-06 | 3.41E-08 | PTK2B | No | No | 0.4533 |
| 35 | rs385893 | 9 | 4763176 | T | C | 0.0153 | 0.0021 | 5.96E-13 | 0.6608 | 4.00E-13 | 6.33E-12 | 8.68E-12 | 7.33E-08 | 4.13E-11 | AK3;RCL1 | No | Yes (AK3) | 0.493 |
| 36 | rs8176645 | 9 | 136149098 | T | A | -0.0145 | 0.0026 | 3.06E-08 | 0.6111 | 2.33E-08 | 1.15E-07 | 5.90E-07 | 1.71E-07 | 3.42E-07 | ABO | No | No | 0.3966 |
| 37 | rs12247015 | 10 | 60145079 | A | G | 0.0337 | 0.0021 | 1.28E-55 | 0.7641 | 9.23E-56 | 2.83E-54 | 1.54E-49 | 5.71E-40 | 1.08E-41 | TFAM | Yes | Yes | 0.4245 |
| 37 | rs4397793 | 10 | 60145597 | G | C | 0.0257 | 0.0022 | 7.38E-33 | 0.8337 | 3.36E-33 | 1.94E-31 | 4.64E-28 | 3.25E-25 | 2.33E-25 | TFAM | Yes | Yes | 0.4185 |
| 38 | rs7896518 | 10 | 65104500 | A | G | -0.0456 | 0.0022 | 9.53E-99 | 0.4708 | 4.20E-97 | 1.77E-95 | 6.82E-91 | 6.34E-63 | 8.29E-74 | JMJD1C | No | No | 0.4354 |
| 39 | rs57066921 | 10 | 102752345 | T | G | -0.1081 | 0.0081 | 8.98E-41 | 0.6303 | 9.91E-40 | 3.76E-40 | 3.94E-36 | 6.73E-29 | 2.31E-29 | TWNK | Yes | Yes | 0.0189 |
| 39 | rs750866 | 10 | 102761458 | A | G | -0.0254 | 0.0026 | 1.27E-22 | 0.9398 | 1.40E-22 | 2.85E-21 | 4.64E-22 | 1.34E-13 | 5.52E-18 | LZTS2 | No | No | 0.1839 |
| 40 | rs34076958 | 10 | 104348848 | T | TTG | 0.0177 | 0.0022 | 2.40E-15 | 0.6309 | 5.84E-15 | 3.21E-15 | 1.48E-13 | 4.08E-10 | 1.83E-14 | SUFU | No | No | 0.3509 |
| 41 | rs139115730 | 11 | 4055492 | T | TTA | 0.0185 | 0.0023 | 5.66E-16 | 0.0413 | 5.17E-16 | 2.04E-16 | 1.13E-17 | 3.87E-10 | 6.70E-09 | STIM1 | No | No | 0.333 |
| 42 | rs11064074 | 12 | 6281039 | C | T | 0.0197 | 0.0021 | 4.87E-20 | 0.1342 | 4.56E-20 | 3.45E-20 | 3.90E-20 | 2.46E-13 | 8.41E-14 | VWF;CD9 | No | No | 0.4294 |
| 43 | rs5012419 | 12 | 19105915 | A | G | 0.0247 | 0.0022 | 2.66E-29 | 0.5733 | 3.68E-29 | 2.91E-27 | 4.12E-29 | 5.73E-18 | 4.23E-22 | CAPZA3;PLEKHA5 | No | No | 0.5915 |
| 44 | rs1127787 | 12 | 27867727 | G | A | -0.0159 | 0.0028 | 1.49E-08 | 0.4664 | 2.33E-08 | 7.98E-08 | 1.37E-08 | 1.93E-03 | 1.57E-08 | MRP535 | No | Yes | 0.173 |
| 45 | rs2015599 | 12 | 29435480 | G | A | 0.0123 | 0.0021 | 6.03E-09 | 0.9942 | 4.94E-09 | 1.06E-08 | 6.42E-08 | 5.26E-09 | 4.27E-06 | LOC100506606 | No | No | 0.4274 |
| 46 | rs12426673 | 12 | 109490296 | G | T | -0.0141 | 0.0021 | 4.03E-11 | 0.2406 | 6.75E-11 | 4.09E-11 | 1.43E-11 | 3.21E-08 | 5.13E-07 | USP30-AS1 | No | Yes | 0.6233 |
| 47 | rs11064881 | 12 | 120146925 | G | A | 0.0255 | 0.004 | 2.58E-10 | 0.231 | 4.27E-10 | 5.39E-10 | 9.77E-10 | 4.72E-07 | 1.70E-08 | CIT | No | No | 0.0676 |
| 48 | rs11553699 | 12 | 122216910 | A | G | 0.0445 | 0.0032 | 1.45E-43 | 0.826 | 6.74E-44 | 2.19E-42 | 3.40E-40 | 5.23E-32 | 7.46E-33 | RHOF;TMEM120B | No | No | 0.1193 |
| 49 | rs1760940 | 14 | 20938251 | A | C | 0.0263 | 0.0024 | 5.20E-27 | 0.5501 | 1.18E-26 | 2.83E-25 | 3.53E-24 | 5.93E-18 | 4.52E-23 | PNP | No | No | 0.2455 |
| 50 | rs72698722 | 14 | 101181881 | C | T | -0.0192 | 0.0027 | 1.25E-12 | 0.9582 | 1.73E-12 | 5.47E-12 | 1.82E-11 | 5.29E-08 | 2.70E-10 | LINC00523;DLK1 | No | No | 0.1938 |
| 51 | rs289713 | 16 | 57006829 | T | A | -0.0149 | 0.0027 | 4.14E-08 | 0.813 | 4.16E-08 | 3.67E-08 | 6.17E-07 | 2.54E-06 | 1.28E-06 | CETP | No | No | 0.7843 |
| 52 | rs3889402 | 17 | 18167397 | G | T | 0.026 | 0.0034 | 2.60E-14 | 0.0428 | 2.86E-14 | 5.63E-14 | 1.25E-13 | 3.80E-07 | 6.69E-13 | MIEF2 | No | Yes | 0.0924 |
| 53 | rs200309755 | 17 | 27845774 | C | T | -0.018 | 0.0022 | 1.91E-16 | 0.5518 | 3.25E-16 | 1.63E-16 | 5.82E-16 | 2.33E-11 | 6.25E-12 | TAOK1 | No | No | 0.5229 |
| 54 | rs8067252 | 17 | 29263700 | C | T | -0.0142 | 0.0026 | 3.58E-08 | 0.5326 | 2.54E-08 | 9.18E-08 | 2.25E-07 | 1.27E-05 | 1.73E-06 | ADAP2 | No | No | 0.2276 |
| 55 | rs17850455 | 17 | 62476451 | C | G | 0.091 | 0.0104 | 1.81E-18 | 0.5969 | 2.96E-18 | 1.48E-19 | 1.13E-19 | 9.28E-11 | 1.08E-13 | POLG2 | Yes | Yes | 0.0129 |
| 56 | rs77261872 | 18 | 42107662 | C | T | 0.0262 | 0.0032 | 2.28E-16 | 0.9872 | 2.68E-16 | 1.46E-16 | 5.61E-17 | 6.14E-10 | 2.27E-11 | LINC01478 | No | No | 0.1431 |
| 57 | rs28665408 | 18 | 67516846 | A | C | 0.0157 | 0.0021 | 2.11E-13 | 0.8681 | 8.92E-14 | 1.46E-12 | 5.39E-12 | 3.58E-11 | 1.58E-09 | DOK6 | No | No | 0.5567 |
| 58 | rs112743753 | 19 | 1000420 | A | G | 0.0116 | 0.0021 | 4.79E-08 | 0.9657 | 4.39E-08 | 2.53E-07 | 2.32E-06 | 1.20E-05 | 6.23E-08 | GRIN3B | No | No | 0.498 |
| 59 | rs11085147 | 19 | 5711930 | C | T | 0.0756 | 0.0036 | 1.54E-95 | 0.3416 | 1.20E-94 | 1.29E-94 | 1.81E-86 | 8.05E-64 | 6.86E-75 | LONP1 | No | Yes | 0.0865 |
| 60 | rs6511720 | 19 | 11202306 | G | T | 0.0192 | 0.0033 | 4.08E-09 | 0.3737 | 8.85E-09 | 3.96E-08 | 4.57E-08 | 2.44E-06 | 1.37E-07 | LDLR | No | No | 0.1103 |
| 61 | rs139891465 | 19 | 17365087 | C | T | 0.0454 | 0.0053 | 1.23E-17 | 0.8923 | 6.74E-18 | 4.68E-18 | 1.17E-16 | 6.73E-15 | 1.15E-10 | USHBP1 | No | No | 0.0398 |
| 61 | rs35586766 | 19 | 17392629 | G | A | 0.0313 | 0.0036 | 4.48E-18 | 0.1762 | 2.67E-32 | 6.14E-32 | 1.08E-30 | 1.25E-22 | 5.08E-24 | ANKLE1 | No | No | 0.1103 |
| 61 | rs56069439 | 19 | 17393925 | C | A | -0.0274 | 0.0023 | 2.30E-32 | 0.7389 | 5.03E-18 | 4.03E-17 | 2.70E-17 | 5.28E-13 | 1.01E-14 | ANKLE1 | No | No | 0.2604 |
| 62 | rs2304128 | 19 | 19746151 | G | T | 0.0226 | 0.0039 | 9.17E-09 | 0.6996 | 1.08E-08 | 1.41E-08 | 2.66E-07 | 4.21E-08 | 9.06E-08 | GMIP | No | No | 0.0805 |
| 63 | rs1065853 | 19 | 45413233 | G | T | 0.0388 | 0.0039 | 1.59E-23 | 0.163 | 4.29E-24 | 1.31E-23 | 2.05E-23 | 3.64E-16 | 5.32E-16 | APOE | No | No | 0.0626 |
| 64 | rs1613662 | 19 | 55536595 | G | A | -0.0167 | 0.0028 | 3.90E-09 | 0.8316 | 3.35E-09 | 1.40E-08 | 7.93E-09 | 4.23E-06 | 8.70E-07 | GP6 | No | No | 0.8549 |
| 65 | rs6074896 | 20 | 1597228 | G | A | -0.0204 | 0.0022 | 1.84E-20 | 0.8619 | 4.53E-21 | 1.84E-19 | 2.45E-19 | 3.22E-09 | 2.37E-19 | SIRPB1 | No | No | 0.6551 |
| 65 | rs156355 | 20 | 1818613 | T | C | 0.0214 | 0.0022 | 6.41E-23 | 0.0768 | 2.31E-22 | 2.42E-22 | 1.18E-20 | 1.78E-15 | 6.40E-17 | LOC100289473;SIRPA | No | No | 0.496 |
| 65 | rs4814776 | 20 | 1921523 | C | A | -0.03 | 0.0023 | 2.00E-40 | 0.3422 | 5.59E-40 | 3.68E-41 | 2.23E-36 | 4.12E-31 | 1.05E-25 | SIRPA | No | No | 0.3469 |
| 66 | rs6105852 | 20 | 17955153 | G | A | 0.0221 | 0.0021 | 1.41E-25 | 0.9089 | 1.11E-25 | 9.43E-27 | 9.72E-27 | 1.27E-14 | 2.65E-19 | MGME1 | Yes | Yes | 0.5179 |
| 66 | rs76599088 | 20 | 17968871 | C | T | 0.0623 | 0.0082 | 3.16E-14 | 0.5873 | 1.07E-14 | 2.84E-13 | 2.49E-11 | 1.48E-07 | 3.01E-15 | MGME1 | Yes | Yes | 0.0139 |
| 67 | rs2245946 | 22 | 43528240 | G | A | 0.0317 | 0.0023 | 9.38E-45 | 0.2517 | 1.53E-44 | 5.83E-45 | 9.18E-39 | 3.09E-31 | 3.08E-36 | MCAT | No | Yes | 0.662 |
| 68 | rs12148 | 22 | 50962208 | T | G | -0.0139 | 0.0022 | 1.32E-10 | 0.9968 | 6.63E-10 | 5.68E-10 | 4.29E-11 | 1.28E-05 | 1.72E-07 | SCO2 | No | Yes | 0.6183 |
| 69 | rs3002416 | X | 39710195 | C | T | -0.0107 | 0.0019 | 6.92E-09 | 0.9789 | 8.60E-09 | 1.12E-08 | 4.59E-08 | 1.50E-06 | 2.12E-06 | MIR1587;BCOR | No | No | 0.4634 |
| 70 | rs200375559 | X | 57434968 | C | CTATT | -0.0133 | 0.0021 | 3.30E-10 | 0.3784 | 1.55E-10 | 1.92E-09 | 1.51E-08 | 4.46E-08 | 8.36E-08 | FAAH2 | No | No | 0.6606 |
| 71 | rs73250616 | X | 99993277 | T | C | 0.0097 | 0.0017 | 2.58E-08 | 0.9324 | 2.86E-08 | 7.93E-08 | 4.52E-07 | 4.27E-06 | 1.02E-07 | SYTL4;CSTF2 | No | No | 0.4922 |
| 72 | rs392020 | X | 118601473 | T | C | -0.0212 | 0.002 | 6.85E-26 | 0.5818 | 6.00E-26 | 1.69E-24 | 5.69E-24 | 6.47E-16 | 1.37E-22 | SLC25A5-AS1 | No | Yes | 0.7454 |

**S2. Table 2**. FINEMAP results for 72 mtDNA-CN loci and 82 independent signals

| | Finemapping Information | | | | |
|---|---|---|---|---|---|
| locus # | 95% Credible Set (SNP Posterior Probability) | # SNPs in Credible Set | Top Finemapping [FM] SNP | Top FM SNP Posterior Probability | Top FM SNP also lead GWAS SNP? | Bayes Factor |
| 1 | rs2977608 (0.999892) | 1 | rs2977608 | 0.999892 | Yes | 7.19 |
| 2 | rs1569419 (0.935346);rs4648452 (0.0268123) | 2 | rs1569419 | 0.935346 | Yes | 4.63 |
| 3 | rs3766744 (0.109848);rs4846082 (0.109848);rs3818157 (0.0733122);rs2236054 (0.0733122);rs4845891 (0.0733122);rs1474868 (0.0733122);rs11121846 (0.0490396);rs2236053 (0.0490396);rs2236055 (0.0490396);rs1548923 (0.0490396);rs1820189 (0.0328776);rs11121845 (0.0328776);rs2180183 (0.0328776);rs3766746 (0.0328776);rs7551289 (0.0220922);rs2273293 (0.0220922);rs6698008 (0.0220922);rs13820190 (0.0148785);rs1040669 (0.0148785);rs6678553 (0.0148785) | 20 | rs3766744 | 0.109848 | Yes | 2.54 |
| 4 | rs139898146 (0.999987) | 1 | rs139898146 | 0.999987 | Yes | 8.36 |
| 5 | rs2274319 (0.0740056);rs1925950 (0.0556045);rs1342442 (0.0556045);rs3790454 (0.0556045);rs6658120 (0.0556045);rs10908505 (0.041865);rs12136856 (0.041865);rs2274317 (0.041865);rs12131289 (0.041865);rs6700679 (0.041865);rs1050316 (0.041865);rs12038396 (0.041865);rs3790459 (0.041865);rs10908504 (0.041865);rs2274320 (0.041865);rs3790457 (0.041865);rs11264486 (0.041865);rs2274316 (0.041865);rs4450010 (0.041865);rs3818463 (0.041865);rs2282286 (0.041865) | 21 | rs2274319 | 0.0740056 | Yes | 2.13 |
| 6 | rs2057301 (0.177559);rs6425520 (0.17611);rs151021588 (0.174181);rs2180747 (0.140774);rs2038480 (0.0540622);rs2038479 (0.0532213);rs6425521 (0.0409674);rs2180748 (0.0399533);rs6662263 (0.0332215);rs10914127 (0.0331587);rs6425522 (0.0330193) | 11 | rs2057301 | 0.177559 | No | 2.54 |
| 7 | rs143989240 (0.3249);rs9425601 (0.200558);rs9425308 (0.124655);rs9425602 (0.12251);rs12901656 (0.053766);rs6691993 (0.0534879);rs6699095 (0.0424655);rs2422141 (0.0218965);rs9425311 (0.00833293) | 9 | rs143989240 | 0.3249 | Yes | 3.24 |
| 8 | rs10749636 (0.794348);rs6659169 (0.0944778);rs11204523 (0.0722311) | 3 | rs10749636 | 0.794348 | Yes | 4.11 |
| 9 | rs116640044 (0.361725);rs540935535 (0.316236);rs115021784 (0.283215) | 3 | rs116640044 | 0.361725 | Yes | 3.11 |
| 10 | rs1653261 (0.186324,rs10543585 (0.119586);rs1726030 (0.0802926);rs2462124 (0.0802104);rs831548 (0.065927);rs2254678 (0.0545873);rs2689773 (0.0541427);rs4852322 (0.0464502);rs2243819 (0.0442816);rs7606862 (0.0304169);rs865551 (0.0250277);rs831522 (0.0225413);rs951409 (0.0197036);rs951408 (0.019699);rs6546881 (0.0196671);rs831524 (0.0195293);rs831523 (0.0133157);rs2694701 (0.0129887);rs831527 (0.0129296);rs2462125 (0.0127278);rs831545 (0.0088729);rs2244751 (0.0086081) | 22 | rs1653261 | 0.186324 | Yes | 2.77 |
| 11 | rs62641680 (1) | 1 | rs62641680 | 1 | Yes | 13.41 |
| 12 | rs74874677 (0.999881) | 1 | rs74874677 | 0.999881 | Yes | 7.30 |
| 13 | rs12052715 (0.193878);rs12692566 (0.0709992);rs72955755 (0.0625224);rs2729708 (0.0535877);rs17231342 (0.0485667);rs56186131 (0.0378116);rs6432570 (0.029505);rs67045814 (0.029505);rs17829052 (0.0230755);rs2063440 (0.0230755);rs6704285 (0.0230755);rs3828323 (0.0230755);rs2063439 (0.0180879);rs2729701 (0.0111896);rs2729702 (0.0111896);rs2955982 (0.0111896);rs2729707 (0.0111896);rs4665122 (0.00936734);rs12052733 (0.00936734);rs1344632 (0.00883092);rs6755783 (0.00883092);rs2556095 (0.00883092);rs2556097 (0.00883092);rs2556098 (0.00883092);rs62175256 (0.00883092);rs2556093 (0.00883092);rs62175254 (0.00883092);rs62175233 (0.00883092);rs12053269 (0.00883092);rs2729703 (0.00883092);rs2556096 (0.00883092);rs55908243 (0.00883092);rs2433689 (0.00883092);rs62175255 (0.00883092);rs2433688 (0.00883092);rs62175257 (0.00883092);rs2556099 (0.00883092);rs12692567 (0.00785181);rs145316090 (0.00785181);rs10674031 (0.00785181);rs62175229 (0.00698521);rs60264981 (0.00698521);rs59785786 (0.00698521);rs62175230 (0.00698521);rs7595388 (0.0055378);rs19857030 (0.0055378);rs2357686 (0.0029738);rs34087946 (0.0029738);rs139012450 (0.0029738);rs12692570 (0.00279713);rs2556094 (0.00279713);rs12221808 (0.00223777);rs3951216 (0.00179424);rs55664214 (0.00179424);rs10186586 (0.00144192);rs11895141 (0.00144192);rs4665125 (0.00144192);rs66814375 (0.00144192);rs11675085 (0.00144192);rs10929957 (0.00144192) | 60 | rs12052715 | 0.193878 | Yes | 13.36 |
| 14 | rs78909033 (0.999997) | 1 | rs78909033 | 0.999997 | Yes | 13.41 |
| 15 | rs13084580 (0.256872);rs7623800 (0.164471);rs60889002 (0.105695);rs35008320 (0.084847);rs59889656 (0.084847);rs2056615 (0.0681735);rs6800630 (0.0548268);rs17791786 (0.0548268);rs13078962 (0.0548268);rs12637881 (0.0548268) | 10 | rs13084580 | 0.256872 | Yes | 3.12 |
| 16 | rs6792510 (0.0914552);rs1352420 (0.0705842);rs4858828 (0.0705842);rs4858798 (0.0705842);rs4858831 (0.0705842);rs12497850 (0.0545887);rs11928552 (0.0545887);rs10865952 (0.0545887);rs12493578 (0.0545887);rs6788650 (0.0545887);rs551116669 (0.0545887);rs7633840 (0.036662);rs6786055 (0.0255662);rs4974084 (0.0199364);rs4858800 (0.00418224);rs6806356 (0.00418224);rs7641050 (0.00418224);rs7610519 (0.00418224);rs10865953 (0.00418224);rs4858830 (0.00418224);rs6442129 (0.00418224);rs6442130 (0.00293607);rs7623023 (0.00293607);rs11920534 (0.00293607);rs7630742 (0.00293607);rs11705888 (0.00293607);rs7431710 (0.00293607);rs7612459 (0.00293607);rs4974082 (0.00293607);rs6791542 (0.00293607);rs7628719 (0.00293607);rs3774628 (0.00293607);rs57423875 (0.00293607);rs9848268 (0.00277777);rs6763557 (0.00277777);rs9882443 (0.00277777);rs13076394 (0.00233246);rs7627404 (0.00233246);rs7430198 (0.00233246);rs1134591 (0.00233246);rs12493289 (0.00233246);rs11705737 (0.00233246);rs6446252 (0.00233246);rs9830636 (0.00233246);rs11130176 (0.00233246);rs9884022 (0.00233246);rs6791234 (0.00233246);rs28642807 (0.00233246);rs4974087 (0.00233246);rs9311433 (0.00233246);rs9311432 (0.00233246);rs4974085 (0.00233246);rs9311434 (0.00233246);rs13064911 (0.00233246);rs6786665 (0.00233246);rs9681717 (0.00233246);rs6446196 (0.00233246);rs7372149 (0.00233246);rs9840684 (0.00233246);rs6446253 (0.00233246);rs13315711 (0.00233246);rs10780033 (0.00233246);rs7643368 (0.00233246);rs7639332 (0.00233246);rs13064883 (0.00233246);rs7430501 (0.00233246);rs11924597 (0.00233246);rs7431857 (0.00233246);rs9755751 (0.00233246);rs7615675 (0.00233246);rs4974079 (0.00233246);rs9311430 (0.00233246);rs7428387 (0.00233246);rs7636057 (0.00233246);rs9754878 (0.00233246);rs9867373 (0.00233246);rs13078308 (0.00233246) | 77 | rs6792510 | 0.0914552 | Yes | 2.12 |
| 17 | rs1354034 (1) | 1 | rs1354034 | 1 | Yes | 13.18 |
| 18 | rs34778241 (0.999639) | 1 | rs34778241 | 0.999639 | Yes | 13.28 |
| 19 | rs9844549 (0.249303);rs13093823 (0.189831);rs13088724 (0.118893);rs12488540 (0.0918667);rs34894010 (0.0510261);rs13089906 (0.039809);rs13072265 (0.0382639);rs13065293 (0.0382639);rs13081984 (0.0227239);rs13072699 (0.0190658);rs2287310 (0.0175538);rs13065234 (0.0157452);rs11400633 (0.0135817);rs13074663 (0.0105252);rs12497946 (0.00755936);rs113885380 (0.00703599);rs145667524 (0.00385725);rs17292964 (0.00385725);rs13083369 (0.00385725);rs13088298 (0.00385725);rs1966448 (0.00385725) | 21 | rs9844549 | 0.249303 | No | 2.86 |
| 20 | rs10013187 (0.045956);rs4270589 (0.0266823);rs58654655 (0.0266823);rs6850747 (0.0266823);rs13128022 (0.0266823);rs4270588 (0.0266823);rs6857428 (0.0266823);rs10516483 (0.0266823);rs12331849 (0.026682);rs7656720 (0.026682);rs5779233 (0.026682);rs7682346 (0.026682);rs7652350 (0.026682);rs9999440 (0.0224864);rs4698839 (0.0224864);rs28396651 (0.0224864);rs2216546 (0.0224864);rs4463068 (0.0224864);rs13144307 (0.0204004);rs7686702 (0.0204004);rs12498921 (0.0204004);rs4530633 (0.0204004);rs34061204 (0.0174265);rs13434472 (0.0174265);rs11736568 (0.0174265);rs4639076 (0.0174265);rs11097760 (0.0174265);rs11097761 (0.0174265);rs3974650 (0.0174265);rs3974649 (0.0174265);rs7679882 (0.0174265);rs6816787 (0.0115331);rs7441177 (0.0120067);rs4607219 (0.0120067);rs11097757 (0.0120067);rs4443287 (0.0120067);rs4610336 (0.0120067);rs4365725 (0.0120067);rs10856963 (0.0120067);rs4426778 (0.0120067);rs4554079 (0.0120067);rs4572884 (0.0105313);rs28625045 (0.0105313);rs10016018 (0.00821229);rs6130052 (0.00821229);rs13108400 (0.00821229);rs4698974 (0.00713085);rs4698972 (0.00551413);rs4235402 (0.00551413);rs4296682 (0.00551413);rs35476121 (0.00309993);rs4345184 (0.0024428);rs13108438 (0.0024428);rs6830569 (0.0024428);rs11942427 (0.0024428);rs13912972 (0.00244239);rs111541987 (0.0019283);rs7662572 (0.0019283);rs11933157 (0.0019283);rs35201947 (0.0019283);rs11940934 (0.0019283);rs13116542 (0.0019283);rs11930376 (0.0019283);rs4698978 (0.0019283);rs4698975 (0.0019283);rs1363067 (0.0019283);rs11934097 (0.0019283);rs34321727 (0.0019283);rs35860418 (0.0019283);rs11939568 (0.0019283) | 70 | rs10013187 | 0.045956 | No | 2.04 |
| 21 | rs7705526 (0.495464);rs2736100 (0.409);rs7726159 (0.0571492) | 3 | rs7705526 | 0.495464 | Yes | 3.45 |
| 22 | rs539151713 (0.0353395);rs1598010 (0.00903469);rs67011705 (0.00903469);rs2338793 (0.00903469);rs10060065 (0.00903469);rs36040109 (0.00903469);rs13181060 (0.0075018);rs10070639 (0.0075018);rs6871969 (0.0075018);rs266423 (0.0075018);rs11953055 (0.0075018);rs10077575 (0.0075018);rs11953668 (0.0075018);rs10059897 (0.0075018);rs35317818 (0.0075018);rs9293450 (0.0075018);rs13184711 (0.0075018);rs3916720 (0.00733003);rs35060674 (0.0067778);rs10039408 (0.00623593);rs34252662 (0.00623593);rs16901045 (0.00623593);rs13171293 (0.00623593);rs13156244 (0.00623593);rs7722344 (0.00623593);rs9293443 (0.00623593);rs11950068 (0.00623593);rs11953876 (0.00623593);rs11950735 (0.00623593);rs6452661 (0.00623593);rs10053857 (0.00623593);rs10036866 (0.00623593);rs6452660 (0.00623593);rs16901131 (0.00623593);rs1317047 (0.00623593);rs7682234 (0.00623593);rs7682233 (0.00623593);rs4254482 (0.00623593);rs9293441 (0.00623593);rs7737643 (0.00623593);rs16901116 (0.00623593);rs2338789 (0.00623593);rs4540158 (0.00623593);rs6037327 (0.00623593);rs10042574 (0.00623593);rs3846645 (0.00623593);rs11960081 (0.00623593);rs10076348 (0.00623593);rs13181437 (0.00623593);rs13157632 (0.00623593);rs6452682 (0.00623593);rs10068605 (0.00623593);rs2338792 (0.00623593);rs11960660 (0.00623593);rs10052974 (0.00623593);rs10068790 (0.00623593);rs10473898 (0.00623593);rs10070210 (0.00623593);rs57787613 (0.00623593);rs10065972 (0.00623593);rs10072874 (0.00623593);rs7702660 (0.00623593);rs13159570 (0.00623593);rs34329209 (0.00623593);rs13177678 (0.00623593);rs6883899 (0.00623593);rs9293437 (0.00623593);rs13190547 (0.00623593);rs10061820 (0.00623593);rs16901102 (0.00623593);rs10041942 (0.00623593);rs13182583 (0.00623593);rs13164198 (0.00623593);rs3916446 (0.00623593);rs9854412 (0.00623593);rs6452670 (0.00623593);rs2294881 (0.00623593);rs4260644 (0.00623593);rs13138966 (0.00623593);rs2338790 (0.00623593);rs34155124 (0.00623593);rs4560525 (0.00623593);rs4288100 (0.00623593);rs34782328 (0.00623593);rs34302030 (0.00623593);rs13174766 (0.00623593);rs61141607 (0.00623593);rs13166571 (0.00623593);rs34066594 (0.00623593);rs13167768 (0.00623593);rs34811011 (0.00623593);rs13155940 (0.00623593);rs11948547 (0.00623593);rs34455289 (0.00623593);rs13187277 (0.00623593);rs55909182 (0.00623593);rs6882092 (0.00567101);rs1951999 (0.00518938);rs3846643 (0.00518938);rs11952024 (0.00518938);rs34664 (0.00518938);rs13168325 (0.00518938);rs6452657 (0.00518938);rs3846639 (0.00518938);rs4624774 (0.00518938);rs3846642 (0.00518938);rs16900954 (0.00518938);rs3846641 (0.00518938);rs34646 (0.00518938);rs11062154 (0.00518938);rs10073222 (0.00518938);rs155429 (0.00518938);rs11952089 (0.00518938);rs34672 (0.00518938);rs11956081 (0.00518938);rs34660 (0.00432329);rs10066265 (0.00432329);rs34653 (0.00432329);rs11954478 (0.00432329);rs155433 (0.00432329);rs155427 (0.00432329);rs155426 (0.00432329);rs34656 (0.00432329);rs35247928 (0.00432329);rs266441 (0.00432329);rs34662 (0.00432329);rs67661762 (0.00432329);rs34663 (0.00432329);rs266430 (0.00432329);rs34955 (0.00432329);rs34665 (0.00432329);rs34661 (0.00432329);rs153320 (0.00432329);rs266437 (0.00432329);rs34667 (0.00432329);rs34658 (0.00432329);rs463250 (0.00432329);rs34655 (0.00432329);rs253169 (0.00432329);rs465388 (0.00432329);rs34657 (0.00432329);rs266425 (0.00432329);rs155431 (0.00432329);rs266426 (0.00432329);rs414249 (0.00432329);rs34670 (0.00432329);rs34029355 (0.00432329);rs76794226 (0.00398252);rs11370979 (0.00398252);rs201731283 (0.00398252);rs13052950 (0.00398252);rs7729120 (0.00398252);rs555315628 (0.00398252);rs7725260 (0.00398252);rs266443 (0.00360574);rs266448 (0.00360574);rs266444 (0.00360574);rs266435 (0.00360574);rs155435 (0.00360574);rs62360681 (0.00360574);rs266446 (0.00360574);rs17606 (0.00360574) | 162 | rs539151713 | 0.0353395 | No | 1.80 |
| 23 | rs114694170 (0.999986) | 1 | rs114694170 | 0.999986 | Yes | 13.07 |
| 24 | rs391629 (0.0248689);rs6905986 (0.0248689);rs969297 (0.0248689);rs212938 (0.0248689);rs212941 (0.0248689);rs926326 (0.0248689);rs212930 (0.0248689);rs212936 (0.0248689);rs301370 (0.0237666);rs301371 (0.0237666);rs301375 (0.0237666);rs214066 (0.0196141);rs34164888 (0.0196141);rs169946 (0.0196141);rs4576241 (0.0196141);rs4711086 (0.0196141);rs214050 (0.0196141);rs214049 (0.0196141);rs214735 (0.0179259);rs301374 (0.0179259);rs301373 (0.0179259);rs301372 (0.0179259);rs301369 (0.0179259);rs1769702 (0.0179259);rs301378 (0.0179259);rs214736 (0.0179259);rs9358861 (0.0154944);rs301379 (0.0153852);rs76903257 (0.00971566);rs35953877 (0.00779691);rs301385 (0.00779691);rs3804110 (0.00639155);rs3804107 (0.00639155);rs403768 (0.00593432);rs391544 (0.00593432);rs7769016 (0.00593432);rs301384 (0.00593432);rs301386 (0.00593432);rs3915740 (0.00593432);rs301381 (0.00593432);rs369799 (0.00593432);rs420098 (0.00593432);rs301387 (0.00593432);rs378713 (0.00593432);rs4559085 (0.00593432);rs34965228 (0.00593432);rs448940 (0.00593432);rs33979638 (0.00593432);rs464096 (0.00593432);rs2762355 (0.00593432);rs10946783 (0.00593432);rs495212 (0.00593432);rs574578 (0.00593432);rs366056 (0.00593432);rs1883259 (0.00523585);rs574515 (0.00452693);rs402965 (0.00452693);rs408734 (0.00452693);rs11409280 (0.00452693);rs489748 (0.00452693);rs7769572 (0.00452693);rs7769551 (0.00452693);rs7751062 (0.00452693);rs4568443 (0.00452693);rs6928493 (0.00452693);rs572657 (0.00452693);rs301380 (0.00452693);rs301382 (0.00452693);rs301383 (0.00452693);rs301390 (0.00452693);rs301389 (0.00452693);rs71310794 (0.00452693);rs11408281 (0.00452693);rs574453 (0.00452693);rs1830798 (0.00452693);rs301391 (0.00452693);rs541918 (0.00452693);rs9461183 (0.00452693);rs301392 (0.00452693);rs4383809 (0.00421035);rs6456685 (0.00421035);rs17266491 (0.00389399);rs2024874 (0.00381505);rs9379775 (0.00381505);rs7771765 (0.00381505);rs7769257 (0.00346114);rs2769701 (0.00346114);rs3804102 (0.00295659);rs7761789 (0.00295659);rs62392334 (0.00283418);rs35785909 (0.00283418);rs7738307 (0.00265228);rs12199135 (0.00265228);rs2762356 (0.00265228);rs4097269 (0.00265228);rs1924471 (0.00265228);rs34946292 (0.00265228);rs406339 (0.00265228);rs406841 (0.00265228);rs392465 (0.00265228);rs7758716 (0.00265228);rs376290 (0.00265228);rs7758527 (0.00265228);rs7738264 (0.00265228);rs67245361 (0.00265228);rs412384 (0.00265228);rs421502 (0.00265228);rs2476824 (0.00265228);rs431039 (0.00265228);rs450681 (0.00265228);rs403453 (0.00265228) | 111 | rs391629 | 0.0248689 | No | 1.95 |
| 25 | rs2844509 (0.177306);rs2071590 (0.0794158);rs2844483 (0.0381266);rs2844484 (0.0381266);rs1121800 (0.0381266);rs2256965 (0.0321027);rs2516390 (0.0299775);rs2516391 (0.0299775);rs928815 (0.0299775);rs2516392 (0.0299775);rs3094595 (0.0299775);rs746868 (0.0299775);rs2523500 (0.0299775);rs2844486 (0.0299775);rs2844485 (0.0299775);rs2516479 (0.0299775);rs2857602 (0.0299775);rs2239704 (0.0299775);rs2857710 (0.0299775);rs3135042 (0.0299775);rs3135043 (0.0299775);rs1266079 (0.0166597);rs929116 (0.0166597);rs2516483 (0.0104631);rs2516488 (0.0104631);rs2516490 (0.0104631);rs2596506 (0.00739838);rs2239705 (0.00622977);rs2854014 (0.00610807);rs2854012 (0.00505161);rs736160 (0.00196569);rs9265909 (0.00158758);rs34911333 (0.00121905);rs1521 (0.00100644);rs3130707 (0.000886577);rs3130629 (0.000846786);rs2844533 (0.000832351) | 37 | rs2844509 | 0.177306 | Yes | 4.23 |
| 26 | rs35372932 (0.197605);rs7745176 (0.131875);rs2004335 (0.131875);rs11756331 (0.0884379);rs11752229 (0.0884379);rs34104805 (0.0848898);rs11754183 (0.0595973);rs11752754 (0.0595973);rs9269234 (0.0220601);rs11756836 (0.0074594);rs10947315 (0.0071369);rs9270170 (0.00707934);rs9270186 (0.00572226);rs13207922 (0.00478847);rs11752186 (0.00417495);rs71542417 (0.00366951);rs11756095 (0.00366951);rs10947314 (0.00361663);rs74729707 (0.00291782);rs10947310 (0.00291058);rs9689437 (0.00278901);rs114258010 (0.00212593);rs9689836 (0.00212593);rs375004483 (0.00155888);rs72838364 (0.00147247);rs2760995 (0.00145197);rs9271459 (0.00087155);rs7749043 (0.000828684);rs707947 (0.00078865);rs7754396 (0.000745847);rs9270636 (0.000734858);rs9270637 (0.000734858);rs9270638 (0.000734858);rs9270549 (0.000701003);rs12722051 (0.000689418);rs72838376 (0.000496899);rs72838461 (0.000479286);rs138669208 (0.000479286);rs114939096 (0.000425144);rs72838360 (0.000425144);rs2647079 (0.000400511);rs2647077 (0.000400511);rs9270667 (0.00039741);rs3667 (0.000380585);rs9270119 (0.000375576);rs9270118 (0.000375576);rs9270120 (0.000375576);rs2647058 (0.000335178);rs9270618 (0.000325005);rs9270612 (0.000325005);rs9270626 (0.000325005);rs72838390 (0.000312262);rs72838249 (0.000312262);rs2656533 (0.000303427);rs184006295 (0.000303427);rs320074763 (0.000297931);rs56892427 (0.000285741);rs28391139 (0.000282465);rs28411161 (0.000282465);rs3021304 (0.000273951);rs9271537 (0.000273951);rs2854011 (0.000268061);rs72838356 (0.000268061);rs72838351 (0.000268061);rs9270631 (0.000266295);rs9270654 (0.000266295);rs9270621 (0.000266295);rs9270655 (0.000266295);rs7757916 (0.000265444);rs28619112 (0.000263016);rs36214157 (0.000263016);rs72838339 (0.000230372);rs3188043 (0.000230372);rs72838349 (0.000230372);rs72838340 (0.000230372);rs72838358 (0.000230372);rs71542421 (0.000228209);rs28644259 (0.000228209);rs34863593 (0.000228209);rs9271364 (0.000220657) | 80 | rs35372932 | 0.197605 | Yes | 3.73 |
| 27 | rs5745582 (0.1642?);rs5745587 (0.1642?);rs75080135 (0.1642?);rs5745568 (0.125782);rs9296095 (0.125782);rs9357161 (0.125782);rs5745564 (0.096435?) | 7 | rs5745582 | 0.1642? | Yes | 2.80 |
| 28 | rs9494142 (0.542825);rs4895441 (0.214436);rs6920211 (0.0472162);rs1331309 (0.0260904);rs9402686 (0.0260904);rs9389269 (0.0194449);rs34164109 (0.00813392);rs9399136 (0.00813392);rs7758845 (0.00813392);rs35959442 (0.00813392);rs9376091 (0.00813392);rs9389268 (0.00813392);rs4895440 (0.00610433);rs9376090 (0.00610433);rs9376092 (0.00610433);rs9402685 (0.00610433);rs9494145 (0.0048029) | 17 | rs9494142 | 0.542825 | No | 3.35 |
| 29 | rs4720497 (0.136822);rs2304694 (0.104343);rs2304693 (0.0734054);rs10951795 (0.0463978);rs12669814 (0.0369585);rs6955183 (0.0369585);rs6943701 (0.0369585);rs12534548 (0.0369585);rs12670202 (0.0369585);rs12672041 (0.0369585);rs41280672 (0.0294771);rs7810661 (0.0294771);rs111656118 (0.0294771);rs73109897 (0.0294771);rs1866583 (0.0294771);rs4724362 (0.0294771);rs7803883 (0.0294771);rs145013958 (0.0294771);rs6965277 (0.0294771);rs6955857 (0.0294771);rs6973982 (0.0294771);rs3735488 (0.00779346);rs111979067 (0.00407684);rs4551245 (0.00329328);rs57170042 (0.00329328);rs73107935 (0.00277509);rs12670523 (0.00266371);rs11763504 (0.00266371);rs12108 (0.00266371);rs1898961 (0.00266371);rs2304692 (0.00266371);rs4724342 (0.00266371);rs11343173 (0.00266371);rs111590366 (0.00266371);rs58841149 (0.00266371) | 35 | rs4720497 | 0.136822 | No | 2.54 |
| 30 | rs4731606 (0.0554474);rs28851188 (0.0554474);rs28854232 (0.0554474);rs2177616 (0.0554474);rs7811282 (0.0422974);rs12671075 (0.0422974);rs1761673 (0.0323392);rs7779873 (0.0323392);rs6467251 (0.0323392);rs819443 (0.0323392);rs1093829 (0.0247816);rs819442 (0.0247816);rs1761645 (0.0247816);rs7793698 (0.0247816);rs1761646 (0.0247816);rs940542 (0.0247816);rs1093834 (0.0247816);rs1093830 (0.0247816);rs11764390 (0.0214531);rs6467258 (0.0190332);rs1320408 (0.0190332);rs7794010 (0.0190332);rs7810280 (0.0146514);rs11464747 (0.011304);rs7786174 (0.011304);rs1404312 (0.011304);rs1851934 (0.011304);rs7790259 (0.011304);rs2366739 (0.00874112);rs10282644 (0.00874112);rs3211849 (0.00677464);rs35369325 (0.00677464);rs1851935 (0.00677464);rs10260012 (0.00677464);rs9649523 (0.00677464);rs9649529 (0.00677464);rs12260020 (0.00677464);rs9649527 (0.00526246);rs3492988 (0.00526246);rs191694 (0.00526246);rs12155030 (0.00526246);rs2366744 (0.00526246);rs1194195 (0.00409709);rs1464793 (0.00409709);rs1608671 (0.00409709);rs1194197 (0.00409709);rs10480808 (0.00409709);rs819444 (0.00409709);rs819456 (0.00409709);rs819457 (0.00409709);rs12706912 (0.00409709);rs1194196 (0.00409709);rs1093835 (0.00409709);rs10246970 (0.00319703);rs1093831 (0.00319703);rs1093833 (0.00319703);rs701269 (0.00319703);rs2781841 (0.00319703);rs10233710 (0.00319703);rs819455 (0.00319703);rs819445 (0.00319703);rs3211839 (0.00250035);rs819460 (0.00250035) | 63 | rs4731606 | 0.0554474 | No | 2.18 |
| 31 | rs445 (0.94887);rs54812782 (0.00395268) | 2 | rs445 | 0.94887 | Yes | 13.14 |
| 32 | rs342292 (0.154141);rs342299 (0.154141);rs342293 (0.154141);rs342296 (0.154141);rs342294 (0.0812802);rs342295 (0.0812802);rs342290 (0.0812802);rs342298 (0.0812802);rs67036916 (0.0429572) | 9 | rs342292 | 0.154141 | No | 2.66 |
| 33 | rs6972244 (0.832639);rs998302 (0.0545145);rs149992290 (0.0125864);rs74750282 (0.0092067);rs77236693 (0.0092067);rs13239317 (0.0068656?);rs12666877 (0.0068656?);rs13236429 (0.0068656?);rs7783637 (0.0052934?);rs56359210 (0.0052934?);rs80273705 (0.0050034?) | 11 | rs6972244 | 0.832639 | Yes | 4.26 |
| 34 | rs7810260 (0.464712);rs8959832 (0.464712);rs35150201 (0.0705727) | 3 | rs7810260 | 0.464712 | No | 3.52 |
| 35 | rs4284061 (0.999475) | 1 | rs4284061 | 0.999475 | Yes | 6.86 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 33 | rs453301 (0.142374);rs4841132 (0.118364);rs4935581 (0.108362);rs9987289 (0.0869917);rs4841133 (0.0641215);rs4240624 (0.0641215);rs983309 (0.0587627);rs439989 (0.0484314);rs393155 (0.0484314);rs6884305 (0.0419263);rs62494425 (0.0419263);rs12547955 (0.0286337);rs2126264 (0.0236566);rs199922514 (0.0159714);rs367543 (0.00624804);rs35420447 (0.00597396);rs36043919 (0.00597396);rs35555151 (0.00510897);rs106380 (0.0048865);rs1461729 (0.0048865);rs2126259 (0.0048865);rs34523847 (0.00461169);rs11783641 (0.00437323);rs35899262 (0.00374689);rs34597048 (0.00330832);rs11781511 (0.00275804);rs75615732 (0.00272327) | 27 | rs453301 | 0.142374 | No | 2.91 |
| 34 | rs2322718 (0.0494943);rs148984304 (0.0494943);rs34947559 (0.0198417);rs17376255 (0.0198417);rs17375610 (0.0198417);rs2163175 (0.0198417);rs5000732 (0.0198417);rs11994437 (0.0198417);rs28739491 (0.0198417);rs28482046 (0.0198417);rs11775519 (0.0198417);rs11777854 (0.0198417);rs11777664 (0.0198417);rs11783759 (0.0198417);rs4732722 (0.0198417);rs17448278 (0.0198417);rs141943259 (0.0198417);rs2163174 (0.0198417);rs2115805 (0.0198417);rs10089637 (0.0198417);rs67432684 (0.0198417);rs2059968 (0.0198417);rs2059969 (0.0198417);rs919492 (0.0198417);rs919491 (0.0198417);rs28496151 (0.0198417);rs1128433 (0.0198417);rs10455510 (0.0198417);rs1045511 (0.0198417);rs1045512 (0.0198417);rs2241650 (0.0198417);rs2241651 (0.0198417);rs3824104 (0.0198417);rs17447958 (0.0198417);rs919495 (0.0198417);rs28671003 (0.0198417);rs17375764 (0.0198417);rs55921260 (0.0198417);rs1429940 (0.0198417);rs17375548 (0.0198417);rs17376241 (0.0198417);rs28612364 (0.0198417);rs4733055 (0.0198417);rs28566530 (0.0198417);rs11775464 (0.0198417) | 45 | rs2322718 | 0.0494943 | Yes | 2.05 |
| 35 | rs385893 (0.980655) | 1 | rs385893 | 0.980655 | Yes | 5.20 |
| 36 | rs8176645 (0.741866);rs657152 (0.0442211);rs544873 (0.0224468);rs543968 (0.0224468);rs643434 (0.0224468);rs613534 (0.0224468);rs494242 (0.0224468);rs644234 (0.0224468);rs8176719 (0.01798);rs9411377 (0.00703985);rs550057 (0.00328198);rs13292932 (0.0029771) | 12 | rs8176645 | 0.741866 | Yes | 13.40 |
| 37 | rs12247015 (1) | 1 | rs12247015 | 1 | Yes | 13.26 |
| 37 | rs4397793 (0.999733) | 1 | rs4397793 | 0.999733 | Yes | 8.25 |
| 38 | rs7080386 (0.173481);rs10740118 (0.173481);rs7075195 (0.0625224);rs10761739 (0.0625224);rs10822159 (0.0625224);rs10822155 (0.0625224);rs7073753 (0.0625224);rs10761741 (0.0625224);rs10761737 (0.0625224);rs7084707 (0.0625224);rs7090111 (0.0625224);rs7098181 (0.0225842);rs7088799 (0.0225842) | 13 | rs7080386 | 0.173481 | No | 3.04 |
| 39 | rs112254994 (0.250699);rs73349121 (0.250142);rs17113613 (0.17238);rs57066921 (0.172128);rs62626301 (0.144022) | 5 | rs112254994 | 0.250699 | No | 2.66 |
| 39 | rs750866 (0.0612128);rs701834 (0.0608764);rs701835 (0.0604091);rs55646950 (0.0602876);rs3740489 (0.0406369);rs11190787 (0.040476);rs7903239 (0.0400618);rs11190786 (0.039955);rs7902510 (0.0398629);rs2295716 (0.0271133);rs67692077 (0.02704);rs12571302 (0.0268856);rs722435 (0.0268082);rs4919511 (0.026735);rs3740488 (0.0267157);rs3740486 (0.0266876);rs111545834 (0.0266612);rs3824783 (0.0266612);rs3740485 (0.0266556);rs3740487 (0.0266532);rs112309064 (0.0266107);rs3740490 (0.0264827);rs41291460 (0.0244959);rs41291462 (0.0242345);rs4919510 (0.0236857);rs67813203 (0.017965);rs735137 (0.0179427);rs2863095 (0.017553);rs1570171 (0.0163162);rs111981433 (0.0162136);rs11255241 (0.0162073) | 31 | rs750866 | 0.0612128 | Yes | 1.95 |
| 40 | rs71474568 (0.447668);rs7475335 (0.216478);rs11591571 (0.216478);rs34377578 (0.0229011);rs112972631 (0.0229011);rs113422568 (0.0147148);rs12762934 (0.0147148) | 7 | rs71474568 | 0.447668 | No | 3.37 |
| 41 | rs145124071 (0.0527457);rs10835540 (0.0372277);rs10835226 (0.0372277);rs4910887 (0.0372277);rs7111104 (0.0372277);rs7932702 (0.0372277);rs4910886 (0.0259145);rs4910889 (0.0259145);rs12290828 (0.0259145);rs35359151 (0.0259145);rs3284450 (0.0259145);rs2268168 (0.0259145);rs111031110 (0.0259145);rs4910890 (0.0259145);rs3817657 (0.0259145);rs7940013 (0.0259145);rs4910891 (0.0259145);rs10835611 (0.0259145);rs10767845 (0.0180766);rs4910904 (0.0180766);rs2056853 (0.0180766);rs9666857 (0.0180766);rs9665829 (0.0180766);rs7949663 (0.0180766);rs9666229 (0.0180766);rs963998 (0.013238);rs4910880 (0.013238);rs11030781 (0.013238);rs10835521 (0.013238);rs10742242 (0.0126354);rs2165728 (0.0126354);rs4910599 (0.0126354);rs4910888 (0.0126354);rs10835684 (0.0126354);rs28419002 (0.0094142);rs10835545 (0.0094142);rs10767811 (0.0094142);rs10835561 (0.0094142);rs7130580 (0.00885024);rs202160322 (0.00719724);rs2119987 (0.00670753);rs10835217 (0.00670753);rs4910877 (0.0047881);rs10767702 (0.00342439);rs10835261 (0.00342439);rs11030599 (0.00342439);rs10835262 (0.00342439);rs4910866 (0.00342439);rs12295138 (0.00342439);rs7932404 (0.00307817);rs4910596 (0.00245371);rs10835556 (0.00245371);rs201179616 (0.00245371);rs10835557 (0.00245371);rs4414209 (0.00245371);rs11459482 (0.00245371);rs10835454 (0.00176515);rs4910875 (0.00176515);rs10835312 (0.00176515);rs10835252 (0.00176515);rs11427975 (0.00176515);rs7120336 (0.00176515);rs10835256 (0.00176515);rs10835311 (0.00176515);rs11030660 (0.00176515);rs11030176 (0.00176515);rs720571 (0.00176515);rs10767715 (0.00176515);rs11030122 (0.00176515);rs10767684 (0.00176515);rs11030631 (0.00176515);rs10767685 (0.00176515);rs67300156 (0.00176515);rs113414602 (0.00176515);rs714380 (0.00176515);rs10835231 (0.00126696);rs4584571 (0.00126696);rs4910869 (0.00126696);rs10835474 (0.00126696);rs6659683 (0.00126696);rs2898947 (0.00126696);rs10835402 (0.00126696);rs10835407 (0.00126696);rs11452047 (0.00126696);rs11030579 (0.00126696);rs11283447 (0.00126696) | 86 | rs145124071 | 0.0527457 | No | 2.40 |
| 42 | rs11064074 (0.629432);rs1362213 (0.166699);rs1362214 (0.0695302);rs7970830 (0.0695302);rs10849414 (0.0124297);rs6489702 (0.0124297) | 6 | rs11064074 | 0.629432 | Yes | 3.66 |
| 43 | rs5012419 (0.206486);rs10841121 (0.124143);rs10841120 (0.124143);rs5012422 (0.124143);rs7484729 (0.124143);rs5012420 (0.0747912);rs5012421 (0.0747912);rs10734683 (0.0747912);rs10841122 (0.0273146) | 9 | rs5012419 | 0.206486 | Yes | 2.82 |
| 44 | rs1127787 (0.0625356);rs7954129 (0.0510991);rs16917499 (0.0510991);rs16915366 (0.0510991);rs71452063 (0.0510991);rs11049110 (0.0510991);rs11049111 (0.0510991);rs7969297 (0.0510991);rs12827496 (0.0418073);rs7973245 (0.0418073);rs16915365 (0.0418073);rs12827504 (0.0418073);rs35366140 (0.0418073);rs11049108 (0.0418073);rs11049102 (0.0418073);rs12146715 (0.0418073);rs7952827 (0.0342487);rs12833700 (0.0280924);rs11049103 (0.0280924);rs12828457 (0.0280924);rs11049104 (0.0230722);rs11049105 (0.0230722);rs11049107 (0.0230722);rs12579998 (0.0189732) | 24 | rs1127787 | 0.0625356 | Yes | 2.20 |
| 45 | rs35024086 (0.234905);rs2015599 (0.234905);rs1006409 (0.177981);rs10771511 (0.177981);rs372063926 (0.039801);rs7969607 (0.0107746);rs34365857 (0.0046388);rs11050173 (0.00448668);rs11050170 (0.00355695);rs10743664 (0.0028257);rs6487809 (0.0022494);rs12231498 (0.0022494);rs16215735 (0.0017526);rs6487807 (0.0017526);rs4331146 (0.0014624);rs7961943 (0.00140422);rs7298031 (0.00140422);rs2016832 (0.00140422);rs7956305 (0.00140422);rs7980218 (0.00140422);rs7959145 (0.00140422);rs10743656 (0.00140422);rs10771514 (0.00140422);rs6487805 (0.00118251);rs7973786 (0.00118251);rs3782509 (0.00118251);rs16934135 (0.00118251);rs7976961 (0.00118251);rs7301188 (0.00118251);rs11050189 (0.00118251);rs7957339 (0.00118251);rs67146951 (0.00118251);rs7954278 (0.00118251);rs10843382 (0.00118251);rs7301398 (0.00118251);rs11050187 (0.00118251);rs7954871 (0.00118251);rs10843384 (0.00118251);rs7312164 (0.00118251);rs7311912 (0.00118251);rs12366527 (0.00118251);rs12366505 (0.00118251);rs11050188 (0.00118251);rs11050181 (0.00118251);rs6487808 (0.00118251);rs3782506 (0.00118251);rs7134796 (0.00114891);rs10771517 (0.00111325);rs57803026 (0.00095797);rs11050193 (0.00095797);rs9738721 (0.00095797);rs10843377 (0.00095797);rs6487804 (0.00092218);rs3782510 (0.00092218);rs12228468 (0.00092218);rs1989478 (0.00092218);rs2288128 (0.00092218) | 57 | rs35024086 | 0.234905 | No | 2.99 |
| 46 | rs4388979 (0.175886);rs10850362 (0.175886);rs3742028 (0.175886);rs12426673 (0.175886);rs10850589 (0.127939);rs10774928 (0.127939) | 6 | rs4388979 | 0.175886 | No | 2.61 |
| 47 | rs11064881 (0.962572) | 1 | rs11064881 | 0.962572 | Yes | 4.77 |
| 48 | rs11553699 (1) | 1 | rs11553699 | 1 | Yes | 13.52 |
| 49 | rs1760940 (0.937336);rs12147688 (0.0626372) | 2 | rs1760940 | 0.937336 | Yes | 4.63 |
| 50 | rs72698722 (0.559123);rs1555405 (0.107431);rs72698721 (0.054031);rs10873520 (0.0336158);rs10151519 (0.0336158);rs1004573 (0.0281997);rs1004574 (0.0281997);rs72698718 (0.0281997);rs59488041 (0.0281997);rs1004572 (0.0281997);rs72698720 (0.0148562);rs4905998 (0.0101036) | 12 | rs72698722 | 0.559123 | Yes | 3.40 |
| 51 | rs289713 (0.409307);rs7499892 (0.140793);rs11076175 (0.11652);rs7203984 (0.0688728);rs8045855 (0.0568884);rs12720922 (0.0389727);rs289714 (0.0224739);rs12720908 (0.0223269);rs1864163 (0.0218205);rs3816117 (0.0167587);rs1800775 (0.0167587);rs11076176 (0.00975738);rs1800777 (0.00830233);rs708272 (0.0067)(078) | 14 | rs289713 | 0.409307 | Yes | 13.31 |
| 52 | rs3889402 (0.0793676);rs142280600 (0.0554614);rs4925157 (0.0507256);rs3088175 (0.0260867);rs2294914 (0.0260867);rs11555081 (0.0260867);rs56230317 (0.0260867);rs56355515 (0.0260867);rs17732686 (0.0260867);rs60435159 (0.0260867);rs56360702 (0.0260867);rs73289931 (0.0260867);rs73289939 (0.0260867);rs58042962 (0.0260867);rs12603700 (0.0260867);rs5746545 (0.0260867);rs73289960 (0.0209363);rs17805907 (0.0209363);rs55845680 (0.0209363);rs55379748 (0.0209363);rs11225885 (0.0168173);rs55912774 (0.0168173);rs56159325 (0.0168173);rs112499673 (0.0168173);rs15210063 (0.0168173);rs78822685 (0.0135204);rs75982164 (0.0135204);rs78521434 (0.0135204);rs113229690 (0.0135204);rs12601400 (0.0135204);rs112192098 (0.0135204);rs56078485 (0.0135204);rs77826279 (0.0135204);rs140286829 (0.0135204);rs79959344 (0.0135204);rs12601687 (0.0135204);rs55940735 (0.0135204);rs74792197 (0.0135204);rs112031627 (0.0135204);rs111436543 (0.0135204);rs17806489 (0.0108792);rs79257979 (0.0108792);rs3794773 (0.0108792);rs75785225 (0.0108792);rs2273026 (0.0108792);rs73289935 (0.0100376) | 46 | rs3889402 | 0.0793676 | Yes | 2.32 |
| 53 | rs1967556 (0.0877994);rs200309755 (0.0251569);rs8065529 (0.0183097);rs56709487 (0.0124434);rs495136 (0.0124434);rs2874402 (0.0124434);rs7211006 (0.0124434);rs6505128 (0.0124434);rs2138855 (0.0124434);rs139745191 (0.0124434);rs7222766 (0.0124434);rs6505121 (0.0124434);rs6505120 (0.0124434);rs80063027 (0.0116377);rs771428 (0.00847575);rs13341521 (0.00847575);rs4794857 (0.00847575);rs7208675 (0.00847575);rs11655000 (0.00847575);rs2586150 (0.00847575);rs2559625 (0.00847575);rs495814 (0.00847575);rs4795509 (0.00847575);rs2681179 (0.00847575);rs6505125 (0.00847575);rs517967 (0.00847575);rs12942804 (0.00847575);rs9914038 (0.00847575);rs11662600 (0.00847575);rs559972 (0.00847575);rs8081111 (0.00847575);rs7210047 (0.00847575);rs480642 (0.00847575);rs6505129 (0.00847575);rs4056946 (0.00847575);rs7208552 (0.00847575);rs8065958 (0.00847575);rs9900280 (0.00847575);rs9897452 (0.00847575);rs9891920 (0.00847575);rs8080967 (0.00847575);rs2320954 (0.00847575);rs502099 (0.00847575);rs8081267 (0.00847575);rs8081858 (0.00847575);rs9896579 (0.00847575);rs477500 (0.00847575);rs2958212 (0.00847575);rs8068172 (0.00847575);rs9916553 (0.00847575);rs12452756 (0.00847575);rs596700 (0.00847575);rs2559627 (0.00847575);rs11652977 (0.00847575);rs5819869 (0.00578633);rs483399 (0.00578633);rs11550930 (0.00578633);rs111885405 (0.00578633);rs708105 (0.00578633);rs1965446 (0.00578633);rs516051 (0.00578633);rs113774008 (0.00578633);rs2559629 (0.00578633);rs16662601 (0.00578633);rs1568373 (0.00578633);rs3110497 (0.00578633);rs7215228 (0.00578633);rs35914520 (0.00578633);rs573377 (0.00578633);rs9914752 (0.00578633);rs1967557 (0.00578633);rs4289056 (0.00578633);rs4794856 (0.00578633);rs11447326 (0.00578633);rs2138853 (0.00578633);rs71138829 (0.00578633);rs10468576 (0.00578633);rs511093 (0.00578633);rs2681178 (0.00578633);rs2586151 (0.00578633);rs526664 (0.00578633);rs533949 (0.00578633);rs8068151 (0.00578633);rs576426 (0.00578633);rs591433 (0.00578633);rs2559620 (0.00578633);rs12943131 (0.00578633);rs499642 (0.00578633);rs4795510 (0.00578633);rs2946741 (0.00578633);rs12936923 (0.00578633);rs4419109 (0.00578633);rs7960227 (0.00578633);rs12947124 (0.00578633);rs9897945 (0.00578633);rs36015749 (0.00578633);rs507577 (0.00395923);rs28414594 (0.00395923);rs514334 (0.00395923);rs9891759 (0.00395923);rs592862 (0.00395923);rs62065240 (0.00395923);rs520641 (0.00395923);rs56974901 (0.00395923);rs4794852 (0.00395923);rs8064434 (0.00395923);rs12945717 (0.00395923);rs2138854 (0.00395923);rs1712652 (0.00395923);rs508694 (0.00395923);rs547306 (0.00395923);rs602056 (0.00395923);rs8068104 (0.00395923);rs2138852 (0.00395923);rs7222804 (0.00395923);rs8081584 (0.00395923);rs9901247 (0.00395923);rs11651639 (0.00395923);rs8076739 (0.00395923);rs605898 (0.00395923);rs59723374 (0.00395923);rs7216918 (0.00395923) | 122 | rs1967556 | 0.0877994 | No | 2.26 |
| 54 | rs8067252 (0.597943);rs75637630 (0.01935);rs76947600 (0.01935);rs62070648 (0.01935);rs62070645 (0.01935);rs117529210 (0.01935);rs79461387 (0.01935);rs62070651 (0.0158648);rs62070652 (0.0158648);rs113934718 (0.0158648);rs111305917 (0.0158648);rs113865680 (0.0151834);rs7223535 (0.0130299);rs62070644 (0.0130299);rs62070643 (0.0130299);rs78961639 (0.00883514);rs200298728 (0.00883514);rs199661266 (0.00883514);rs35958868 (0.00729424);rs35840638 (0.00729424);rs11867227 (0.00603253);rs8070670 (0.00587189);rs62068770 (0.00499773);rs62070628 (0.00499773);rs4506963 (0.00499773);rs62068782 (0.00470357);rs62070637 (0.00429512);rs62070661 (0.00429512);rs13051880 (0.00363482);rs6505216 (0.00354708);rs6505211 (0.00310051);rs62070626 (0.00310051);rs2079670 (0.00310051);rs62070630 (0.00310051);rs8069884 (0.00310051);rs7221817 (0.00310051);rs4384651 (0.00310051);rs75025587 (0.00308402);rs9900596 (0.00281402);rs113651152 (0.00281402) | 40 | rs8067252 | 0.597943 | Yes | 13.22 |
| 55 | rs17850455 (1) | 1 | rs17850455 | 1 | Yes | 10.23 |
| 56 | rs11082396 (0.165795);rs141236451 (0.0571819);rs77261872 (0.0544208);rs9945167 (0.0439348);rs9945359 (0.0439348);rs9944794 (0.0439348);rs74260365 (0.0439348);rs75763843 (0.0337918);rs75957619 (0.0337918);rs9964138 (0.0260174);rs9951731 (0.0260174);rs9963592 (0.0260174);rs79824385 (0.0260174);rs79010100 (0.0260174);rs117941510 (0.0253888);rs16978036 (0.0253888);rs17841683 (0.0253888);rs28497983 (0.0200525);rs74997723 (0.0200525);rs138805966 (0.0197295);rs9947269 (0.0154713);rs75354229 (0.0153466);rs11082383 (0.0119491);rs12454829 (0.0119491);rs12456054 (0.0119491);rs12606438 (0.00931276);rs11082384 (0.00931276);rs4586544 (0.00931276);rs11082385 (0.00931276);rs80277818 (0.00931276);rs869899 (0.00726519);rs28547892 (0.00726519);rs76537328 (0.00726519);rs10502831 (0.00726519);rs28782011 (0.00726519);rs12456660 (0.00726519);rs11874778 (0.00726519) | 37 | rs11082396 | 0.165795 | No | 2.70 |
| 57 | rs28665408 (0.57391);rs188229 (0.0595624);rs1499414 (0.0595624);rs1788229 (0.025617);rs72739671 (0.0183409);rs34875178 (0.0183409);rs1790931 (0.0183409);rs763361 (0.013161);rs4891780 (0.013161);rs1788105 (0.013161);rs3018275 (0.013161);rs1610554 (0.00946547);rs13024512 (0.00946547);rs1790932 (0.00946547);rs34086233 (0.00946547);rs1790961 (0.00946547);rs13381239 (0.00946547);rs9950174 (0.00946547);rs1808094 (0.00682305);rs1790576 (0.00682305);rs1790947 (0.00682305);rs7244202 (0.00682305);rs1788232 (0.00682305);rs1615504 (0.00682305);rs8099412 (0.00682305);rs1790963 (0.00492946);rs1790946 (0.00492946) | 27 | rs28665408 | 0.57391 | Yes | 3.49 |
| 58 | rs112743753 (0.0834059);rs12986002 (0.0834059);rs34418036 (0.0642043);rs35581531 (0.0642043);rs62134211 (0.0642043);rs35873901 (0.0642043);rs56119352 (0.0527062);rs11669351 (0.0495354);rs2075606 (0.0424423);rs12232842 (0.0390666);rs112236900 (0.0383046);rs2301813 (0.037099);rs71335282 (0.0296872);rs111287086 (0.0243874);rs11084890 (0.0141434);rs741765 (0.0141434);rs3766440 (0.0141434);rs741764 (0.0141434);rs2242436 (0.013172);rs12976079 (0.00951805);rs11879091 (0.0085886);rs4312418 (0.0084015);rs2074455 (0.0084015);rs6510985 (0.00794853);rs1683584 (0.00794853);rs1637996 (0.00649339);rs2238584 (0.00649339);rs4807313 (0.00649339);rs758218 (0.00649339);rs886992 (0.00649339);rs2238585 (0.00649339);rs8107531 (0.00649339);rs8105235 (0.00649339);rs4807312 (0.00649339);rs1029690 (0.00531313);rs1548666 (0.00531313);rs11878920 (0.00419315);rs6844 (0.00368001);rs11669794 (0.00331676);rs11670014 (0.00331676);rs7258402 (0.00331676);rs11670089 (0.00331676);rs8113305 (0.00306895) | 43 | rs112743753 | 0.0834059 | Yes | 2.75 |
| 59 | rs11085147 (1) | 1 | rs11085147 | 1 | Yes | 11.96 |
| 60 | rs142158911 (0.0490098);rs6511720 (0.0490098);rs57217136 (0.0345215);rs14846969 (0.0345215);rs17248727 (0.0345215);rs73015024 (0.0244057);rs138294113 (0.0244057);rs142130958 (0.0244057);rs148895583 (0.0244057);rs61194703 (0.0244057);rs73015011 (0.0244057);rs10412048 (0.0244057);rs73015020 (0.0244057);rs12151108 (0.0244057);rs14826254 (0.0244057);rs10402112 (0.0244057);rs145960625 (0.021625);rs143020224 (0.0205489);rs114821903 (0.0205489);rs147985405 (0.0205489);rs115594766 (0.0205489);rs55997232 (0.0205489);rs112107114 (0.0205489);rs138175288 (0.0205489);rs55791371 (0.0205489);rs111989435 (0.0205489);rs113722226 (0.0205489);rs112032422 (0.0205489);rs375846700 (0.0205489);rs73015013 (0.0205489);rs77265569 (0.0183935);rs112374545 (0.0173175);rs112898275 (0.0173175);rs56289821 (0.0173175);rs112552009 (0.0173175);rs112736558 (0.0173175);rs77140532 (0.0173175);rs139853365 (0.0159397);rs73015021 (0.0146076);rs15113958 (0.0146076);rs56125973 (0.0146076);rs145329186 (0.0123331) | 42 | rs142158911 | 0.0490098 | No | 2.05 |
| 61 | rs139891465 (0.996324) | 1 | rs139891465 | 0.996324 | No | 5.86 |
| 61 | rs4808075 (0.105709);rs10419397 (0.104948);rs56069439 (0.104509);rs13343778 (0.0980371);rs55924783 (0.0738152);rs4808616 (0.0733522);rs113299211 (0.0729241);rs67397200 (0.0725952);rs111961716 (0.0709743);rs4808076 (0.0702698);rs200337891 (0.0514705);rs28473003 (0.0502189);rs61494113 (0.0480526) | 13 | rs4808075 | 0.105709 | No | 2.41 |
| 61 | rs35586766 (0.677972);rs6512183 (0.11045);rs10420922 (0.0912805);rs35006574 (0.0849748) | 4 | rs35586766 | 0.677972 | No | 3.89 |
| 62 | rs2304128 (0.165305);rs58847337 (0.0755947);rs73004962 (0.0755947);rs7504426 (0.0651692);rs16996185 (0.0651692);rs12610185 (0.0651692);rs12610191 (0.0651692);rs10500212 (0.0651692);rs73004966 (0.0562203);rs73004975 (0.0419273);rs73004959 (0.0362451);rs150824230 (0.0313547);rs12608729 (0.0271428);rs73004926 (0.0254005);rs73004933 (0.0220916);rs141756246 (0.0220916);rs73004951 (0.0203828);rs73004967 (0.018664);rs17217098 (0.018664) | 19 | rs2304128 | 0.165305 | Yes | 2.67 |
| 63 | rs1065853 (0.780059);rs7412 (0.219939) | 2 | rs1065853 | 0.780059 | Yes | 3.94 |
| 64 | rs892090 (0.234067);rs1613662 (0.234067);rs1654425 (0.153324);rs1269736 (0.12433);rs61658003 (0.0856496);rs1671152 (0.0102461);rs4020012 (0.00951049);rs1654444 (0.00755425);rs1671219 (0.00755425);rs1654442 (0.00755425);rs1626871 (0.00755425);rs16671220 (0.00755425);rs1671226 (0.00712321);rs775824 (0.00624392);rs1654440 (0.00624392);rs8100773 (0.00546473);rs8099942 (0.00546473);rs35266089 (0.00501394);rs3060075 (0.00409288);rs1671217 (0.00409288);rs1654505 (0.00406343);rs4806635 (0.00355284);rs2569513 (0.00355284);rs1654432 (0.00295159);rs1654433 (0.00245521);rs1654447 (0.00239388) | 26 | rs892090 | 0.234067 | No | 3.12 |
| 65 | rs6074896 (0.999241) | 1 | rs6074896 | 0.999241 | No | 6.59 |
| 65 | rs156355 (0.943727);rs156333 (0.0064512) | 2 | rs156355 | 0.943727 | Yes | 4.66 |
| 65 | rs4814778 (0.462974);rs8136489 (0.282138);rs8112072 (0.252192) | 3 | rs4814778 | 0.462974 | No | 3.38 |
| 66 | rs6105852 (0.120099);rs6105851 (0.119793);rs8121099 (0.119786);rs6111779 (0.0682359);rs754169 (0.0565721);rs8121298 (0.0358655);rs7267254 (0.0358036);rs11696758 (0.0357727);rs751902 (0.0357704);rs984853 (0.0350043);rs11697739 (0.0349676);rs3790301 (0.0349008);rs6111753 (0.0348385);rs6136235 (0.0348084);rs9100900 (0.0347891);rs982121 (0.0344711);rs982120 (0.0344689);rs16978653 (0.0341981);rs6136244 (0.028315) | 19 | rs6105852 | 0.120099 | Yes | 2.50 |
| 66 | rs76599088 (0.687836);rs75469378 (0.206419);rs186228279 (0.0880411) | 3 | rs76599088 | 0.687836 | Yes | 3.73 |
| 67 | rs13815 (0.137417);rs2249946 (0.137417);rs2245947 (0.137417);rs5759176 (0.137417);rs737802 (0.0755889);rs2076079 (0.0755889);rs2066910 (0.0755889);rs6003067 (0.0755889);rs2235317 (0.0416577);rs2235316 (0.0416577);rs7402 (0.0416577) | 11 | rs13815 | 0.137417 | No | 2.64 |
| 68 | rs140522 (0.316625);rs12148 (0.132648);rs131814 (0.0996661);rs2782 (0.0996661);rs131808 (0.0996661);rs470119 (0.0750394);rs140523 (0.0750394);rs131801 (0.0750394) | 8 | rs140522 | 0.316625 | No | 3.30 |
| 69 | rs3002417 (0.610746);rs3002416 (0.365261) | 2 | rs3002417 | 0.610746 | No | 3.32 |

**S2. Table 3**. Overlap between mtDNA-CN loci and MT-eQTLs (Ali *et al.*, 2019)

| rsid | chr | pos | ref | alt | Nuclear Gene | Nuclear Gene Annotation | mtDNA-CN GWAS Beta | mtDNA-CN GWAS P-value | MT-eQTL Tissue | MT-eQTL Gene | MT-eQTL GWAS Beta | MT-eQTL GWAS P-value | Concordant sign between mtDNA-CN and MT-eQTL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs11085147 | 19 | 5711930 | C | T | *LONP1* | NM_001276479:exon5:c.G530A:p.R177Q | 0.0756 | 1.54E-95 | Whole Blood | *MTCO2* | NA | NA | NA |
| | | | | | | | | | | *MTCYB* | NA | NA | NA |
| | | | | | | | | | | *MTCO2* | -0.028 | 7.53E-31 | No |
| | | | | | | | | | | *MTCO3* | -0.028 | 1.23E-25 | No |
| | | | | | | | | | | *MTND1* | 0.015 | 2.34E-17 | Yes |
| | | | | | | | | | Skin (not sun exposed) | *MTCO2* | NA | NA | NA |
| | | | | | | | | | | *MTCO3* | -0.027 | 2.84E-12 | No |
| | | | | | | | | | Lymphoblastoid cell lines | *MTND4L* | NA | NA | NA |
| | | | | | | | | | | *MTRNR2* | NA | NA | NA |
| rs2304693 | 7 | 45148667 | G | A | *TBRG4* | NM_199122:exon2:c.C170T:p.P57L | 0.0181 | 5.11E-11 | | *MTCO2* | NA | NA | NA |
| | | | | | | | | | | *MTCO3* | NA | NA | NA |
| | | | | | | | | | | *MTND1* | NA | NA | NA |
| | | | | | | | | | | *MTND4* | -0.029 | 2.35E-18 | No |
| | | | | | | | | | Subcutaneous adipose | *MTCO3* | NA | NA | NA |
| | | | | | | | | | | *MTND1* | NA | NA | NA |
| | | | | | | | | | | *MTND5* | NA | NA | NA |
| | | | | | | | | | | *MTND6* | NA | NA | NA |
| rs1127787 | 12 | 27867727 | G | A | *MRPS35* | NM_021821:exon2:c.G127A:p.G43R | -0.0159 | 1.49E-08 | Subcutaneous Adipose | *MTND1* | -0.032 | 2.48E-23 | Yes |
| | | | | | | | | | | *MTND4* | -0.022 | 2.20E-12 | Yes |
| | | | | | | | | | Skin (not sun exposed) | *MTND1* | -0.022 | 7.72E-12 | Yes |
| rs62165226 | 2 | 55878339 | T | C | *PNPT1* | intronic | -0.0085 | 5.47E-05 | Whole Blood | *MTRNR2* | NA | NA | NA |
| | | | | | | | | | | *MTND5* | NA | NA | NA |
| | | | | | | | | | | *MTND6* | NA | NA | NA |
| | | | | | | | | | Lymphoblastoid cell lines | *MTCO3* | -0.018 | 2.79E-08 | Yes |
| rs10205130 | 2 | 44151572 | G | C | *LRPPRC* | intronic | -0.0084 | 1.17E-04 | Lymphoblastoid cell lines | *MTCO2* | NA | NA | NA |
| | | | | | | | | | | *MTND1* | 0.023 | 1.23E-10 | No |

**S2. Table 4**. DEPICT gene prioritization results

| SNPs in Locus | # Genes in Locus | Locus Coordinates | GWAS P-value | Ensembl Gene ID | Gene Symbol | Included in GeneMANIA Analysis | DEPICT P-value (Nominal) | DEPICT FDR | Gene Closest to Lead SNP | Top cis-eQTL (Westra et al, 2014; Nature Genetics) |
|---|---|---|---|---|---|---|---|---|---|---|
| rs55955211 | 4 | chr22:50528308-50656045 | 2.53E-08 | ENSG00000073169 | SELO | Yes | 3.07E-08 | <=0.01 | TRUE | rs2272847 |
| rs465299 | 1 | chr9:4860454-4885917 | 1.31E-08 | ENSG00000243754 | RP11-125K10.4 | No | 1.29E-07 | <=0.01 | TRUE | - |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000135632 | SMYD5 | Yes | 2.81E-06 | <=0.01 | FALSE | rs2421547 |
| rs1760940 | 1 | chr14:20937113-20945253 | 5.20E-27 | ENSG00000198805 | PNP | Yes | 9.68E-06 | <=0.01 | TRUE | rs1713459 |
| rs11697158;rs6105852;rs76599088 | 4 | chr20:17922241-18039832 | 1.41E-25 | ENSG00000089006 | SNX5 | Yes | 1.09E-05 | <=0.01 | TRUE | rs6132012 |
| rs55955211 | 4 | chr22:50528308-50656045 | 2.53E-08 | ENSG00000170638 | TRABD | Yes | 1.36E-05 | <=0.01 | FALSE | rs732710 |
| rs12780745;rs1408343 | 6 | chr10:102672720-102790890 | 6.26E-28 | ENSG00000055950 | MRPL43 | Yes | 1.48E-05 | <=0.01 | FALSE | rs2863095 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000148840 | PPRC1 | Yes | 2.21E-05 | <=0.01 | FALSE | rs10736156 |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000144048 | DUSP11 | Yes | 2.30E-05 | <=0.01 | FALSE | - |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000178537 | SLC25A20 | Yes | 4.76E-05 | <=0.01 | FALSE | - |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000141994 | DUS3L | Yes | 4.88E-05 | <=0.01 | FALSE | - |
| rs2304693 | 5 | chr7:45002261-45151646 | 5.11E-11 | ENSG00000136270 | TBRG4 | Yes | 5.04E-05 | <=0.01 | TRUE | rs3757573;rs4724362;rs2289373 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000166197 | NOLC1 | Yes | 8.17E-05 | <=0.01 | FALSE | - |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000166197 | NOLC1 | Yes | 8.17E-05 | <=0.01 | FALSE | - |
| rs12780745;rs1408343 | 6 | chr10:102672720-102790890 | 6.26E-28 | ENSG00000107815 | C10orf2 | Yes | 8.37E-05 | <=0.01 | FALSE | rs927302 |
| rs12451698;rs3889402 | 10 | chr17:18012020-18317694 | 2.60E-14 | ENSG00000177731 | FLII | Yes | 1.02E-04 | <=0.01 | FALSE | rs8065874 |
| rs10407593;rs11085147;rs9636179 | 7 | chr19:5558178-5720463 | 1.54E-95 | ENSG00000196365 | LONP1 | Yes | 1.59E-04 | <=0.01 | TRUE | rs7256359 |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000172053 | QARS | Yes | 2.57E-04 | <=0.01 | FALSE | rs11706052 |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000105726 | ATP13A1 | Yes | 2.75E-04 | <=0.01 | FALSE | rs2304130 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000077150 | NFKB2 | Yes | 3.09E-04 | <=0.01 | TRUE | - |
| rs12451698;rs3889402 | 10 | chr17:18012020-18317694 | 2.60E-14 | ENSG00000176974 | SHMT1 | Yes | 3.41E-04 | <=0.01 | FALSE | rs2168781 |
| rs139891465;rs35586766;rs56069439;rs7254318 | 8 | chr19:17360838-17453539 | 2.30E-32 | ENSG00000130312 | MRPL34 | Yes | 3.84E-04 | <=0.01 | FALSE | rs8100448 |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000230002 | ALMS1-IT1 | No | 4.28E-04 | <=0.01 | FALSE | - |
| rs111870993;rs117263028;rs117821325;rs145707459;rs56052501;rs7896518;rs7916282 | 3 | chr10:64893050-65384883 | 9.53E-99 | ENSG00000148572 | NRBF2 | Yes | 5.26E-04 | <=0.01 | TRUE | - |
| rs148308452 | 10 | chr12:122277433-122907179 | 3.70E-09 | ENSG00000175727 | MLXIP | Yes | 6.21E-04 | <=0.01 | FALSE | - |
| rs12148 | 4 | chr22:50946645-50971009 | 1.32E-10 | ENSG00000025708 | TYMP | Yes | 7.13E-04 | <=0.01 | FALSE | rs140522;rs140522;rs2341010 |
| rs11553699;rs9804982 | 4 | chr12:122064455-122232261 | 1.45E-43 | ENSG00000139725 | RHOF | Yes | 8.25E-04 | <=0.01 | TRUE | rs11043194 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000166169 | POLL | Yes | 8.40E-04 | <=0.01 | FALSE | - |
| rs342293 | 1 | chr7:106297211-106301442 | 1.05E-40 | ENSG00000253276 | C7orf74 | Yes | 9.15E-04 | <=0.01 | TRUE | - |
| rs13084580 | 6 | chr3:38887260-39234087 | 2.19E-13 | ENSG00000144655 | CSRNP1 | Yes | 1.03E-03 | <=0.01 | TRUE | rs3732383 |
| rs1127787 | 1 | chr12:27863706-27909228 | 1.49E-08 | ENSG00000061794 | MRPS35 | Yes | 1.04E-03 | <=0.01 | TRUE | rs10771360 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000166189 | HPS6 | Yes | 1.07E-03 | <=0.01 | FALSE | rs17771448 |
| rs13088724 | 3 | chr3:179040779-179169378 | 7.70E-12 | ENSG00000121864 | ZNF639 | Yes | 1.08E-03 | <=0.01 | FALSE | - |
| rs8067252 | 3 | chr17:29158988-29286340 | 3.58E-08 | ENSG00000172171 | C17orf42 | Yes | 1.10E-03 | <=0.01 | FALSE | - |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000089639 | GMIP | Yes | 1.22E-03 | <=0.01 | FALSE | rs1476459 |
| rs2245946 | 2 | chr22:43506754-43539400 | 9.38E-45 | ENSG00000100294 | MCAT | Yes | 1.29E-03 | <=0.01 | TRUE | - |
| rs139891465;rs35586766;rs56069439;rs7254318 | 8 | chr19:17360838-17453539 | 2.30E-32 | ENSG00000130299 | GTPBP3 | Yes | 1.31E-03 | <=0.01 | FALSE | rs3826700;rs7247558 |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000129933 | MAU2 | Yes | 1.34E-03 | <=0.01 | FALSE | rs2301668 |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000177479 | ARIH2 | Yes | 1.54E-03 | <0.05 | FALSE | rs7628719 |
| rs7223593 | 1 | chr17:17991239-18011285 | 4.79E-08 | ENSG00000108591 | DRG2 | Yes | 1.73E-03 | <0.05 | TRUE | rs854791 |
| rs2304693 | 5 | chr7:45002261-45151646 | 5.11E-11 | ENSG00000232956 | SNHG15 | No | 1.87E-03 | <0.05 | FALSE | - |
| rs112743753 | 2 | chr19:984328-1009731 | 4.79E-08 | ENSG00000065268 | WDR18 | Yes | 1.98E-03 | <0.05 | FALSE | rs4806884 |
| rs34243225;rs6959832 | 3 | chr7:135046547-135378166 | 2.10E-23 | ENSG00000155561 | NUP205 | Yes | 2.00E-03 | <0.05 | TRUE | rs13241136 |
| rs12247015;rs186832534;rs2790203 | 4 | chr10:59951278-60158981 | 1.28E-55 | ENSG00000108064 | TFAM | Yes | 2.21E-03 | <0.05 | TRUE | rs10826176 |
| rs10749636 | 1 | chr1:248020501-248041507 | 5.43E-10 | ENSG00000162722 | TRIM58 | Yes | 2.26E-03 | <0.05 | TRUE | rs10788730 |
| rs12426673 | 3 | chr12:109460894-109531436 | 4.03E-11 | ENSG00000189046 | ALKBH2 | Yes | 2.44E-03 | <0.05 | FALSE | rs246085 |
| rs12780745;rs1408343 | 6 | chr10:102672720-102790890 | 6.26E-28 | ENSG00000119906 | FAM178A | Yes | 2.46E-03 | <0.05 | TRUE | rs10883567 |
| rs17850455 | 1 | chr17:62473904-62493184 | 1.81E-18 | ENSG00000256525 | POLG2 | Yes | 2.51E-03 | <0.05 | TRUE | - |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000185909 | KLHDC8B | Yes | 2.53E-03 | <0.05 | FALSE | rs11706189 |
| rs139891465;rs35586766;rs56069439;rs7254318 | 8 | chr19:17360838-17453539 | 2.30E-32 | ENSG00000160117 | ANKLE1 | Yes | 2.56E-03 | <0.05 | TRUE | rs9676419 |
| rs11553699;rs9804982 | 4 | chr12:122064455-122232261 | 1.45E-43 | ENSG00000182500 | ORAI1 | Yes | 3.06E-03 | <0.05 | FALSE | rs12308869 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000107833 | NPM3 | Yes | 3.18E-03 | <0.05 | FALSE | rs2305191 |
| rs10407593;rs11085147;rs9636179 | 7 | chr19:5558178-5720463 | 1.54E-95 | ENSG00000174917 | C19orf70 | Yes | 3.34E-03 | <0.05 | TRUE | - |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs78909033 | 2 | chr2:241505221-241557122 | 1.00E-11 | ENSG00000142327 | RNPEPL1 | Yes | 3.54E-03 | <0.05 | FALSE | rs4676430 |
| rs11553699;rs9804982 | 4 | chr12:122064455-122232261 | 1.45E-43 | ENSG00000188735 | TMEM120B | Yes | 3.69E-03 | <0.05 | TRUE | - |
| rs11697158;rs6105852;rs76599088 | 4 | chr20:17922241-18039832 | 1.41E-25 | ENSG00000218902 | PTMAP3 | No | 4.34E-03 | <0.05 | FALSE | - |
| rs10407593;rs11085147;rs9636179 | 7 | chr19:5558178-5720463 | 1.54E-95 | ENSG00000130254 | SAFB2 | Yes | 4.38E-03 | <0.05 | FALSE | rs639858;rs708691 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000107874 | CUEDC2 | Yes | 4.54E-03 | <0.05 | FALSE | rs11191274 |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000172046 | USP19 | Yes | 4.80E-03 | <0.05 | FALSE | - |
| rs10407593;rs11085147;rs9636179 | 7 | chr19:5558178-5720463 | 1.54E-95 | ENSG00000160633 | SAFB | Yes | 4.82E-03 | <0.05 | FALSE | rs8102642 |
| rs3766744 | 2 | chr1:11994262-12073571 | 2.91E-17 | ENSG00000116688 | MFN2 | Yes | 5.06E-03 | <0.05 | TRUE | rs873458 |
| rs1613662 | 3 | chr19:55476438-55580914 | 3.90E-09 | ENSG00000088053 | GP6 | Yes | 5.15E-03 | <0.05 | TRUE | rs17836542 |
| rs13084580 | 6 | chr3:38887260-39234087 | 2.19E-13 | ENSG00000114742 | WDR48 | Yes | 5.33E-03 | <0.05 | FALSE | rs3736573 |
| rs11064881 | 3 | chr12:119825792-120315095 | 2.58E-10 | ENSG00000111725 | PRKAB1 | Yes | 5.84E-03 | <0.05 | FALSE | rs11064881 |
| rs34243225;rs6959832 | 3 | chr7:135046547-135378166 | 2.10E-23 | ENSG00000243317 | C7orf73 | Yes | 5.86E-03 | <0.05 | FALSE | - |
| rs34594414 | 1 | chr18:67528097-67624160 | 1.07E-12 | ENSG00000150637 | CD226 | Yes | 6.27E-03 | <0.05 | TRUE | rs763361 |
| rs148308452 | 10 | chr12:122277433-122907179 | 3.70E-09 | ENSG00000130779 | CLIP1 | Yes | 6.34E-03 | <0.05 | FALSE | - |
| rs12451698;rs3889402 | 10 | chr17:18012020-18317694 | 2.60E-14 | ENSG00000177302 | TOP3A | Yes | 6.38E-03 | <0.05 | FALSE | rs1563632 |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000167491 | GATAD2A | Yes | 6.42E-03 | <0.05 | FALSE | rs6909;rs7259773 |
| rs11591571 | 4 | chr10:104221149-104418164 | 3.77E-15 | ENSG00000138107 | ACTR1A | Yes | 7.28E-03 | <0.05 | FALSE | rs2902544 |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000114316 | USP4 | Yes | 8.00E-03 | <0.05 | FALSE | rs11713297;rs3774800 |
| rs11697158;rs6105852;rs76599088 | 4 | chr20:17922241-18039832 | 1.41E-25 | ENSG00000125871 | C20orf72 | Yes | 9.05E-03 | <0.05 | TRUE | rs8120495 |
| rs11764390 | 1 | chr7:79998891-80308593 | 1.55E-08 | ENSG00000135218 | CD36 | Yes | 9.72E-03 | <0.05 | FALSE | rs3211958;rs2272353 |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000178035 | IMPDH2 | Yes | 9.90E-03 | <0.05 | FALSE | - |
| rs12148 | 4 | chr22:50946645-50971009 | 1.32E-10 | ENSG00000025770 | NCAPH2 | Yes | 0.01 | <0.05 | FALSE | rs140524;rs5770769 |
| rs12247015;rs186832534;rs2790203 | 4 | chr10:59951278-60158981 | 1.28E-55 | ENSG00000072401 | UBE2D1 | Yes | 0.01 | <0.05 | TRUE | rs10826174 |
| rs2245946 | 2 | chr22:43506754-43539400 | 9.38E-45 | ENSG00000100290 | BIK | Yes | 0.01 | <0.05 | FALSE | rs5751435 |
| rs4698839 | 1 | chr4:102332443-102995969 | 9.47E-09 | ENSG00000153064 | BANK1 | Yes | 0.01 | <0.05 | TRUE | rs17031974 |
| rs139891465;rs35586766;rs56069439;rs7254318 | 8 | chr19:17360838-17453539 | 2.30E-32 | ENSG00000105393 | BABAM1 | Yes | 0.01 | <0.05 | FALSE | rs7246262;rs891017 |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000174886 | NDUFA11 | Yes | 0.01 | <0.05 | FALSE | - |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000198728 | LDB1 | Yes | 0.01 | <0.05 | FALSE | - |
| rs10835540 | 2 | chr11:3875757-4160106 | 5.74E-16 | ENSG00000167325 | RRM1 | Yes | 0.01 | <0.05 | TRUE | rs12806698 |
| rs8067252 | 3 | chr17:29158988-29286340 | 3.58E-08 | ENSG00000176208 | ATAD5 | Yes | 0.01 | <0.05 | FALSE | - |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000135624 | CCT7 | Yes | 0.01 | <0.05 | FALSE | rs12464589;rs12104774 |
| rs12148 | 4 | chr22:50946645-50971009 | 1.32E-10 | ENSG00000130489 | SCO2 | Yes | 0.01 | <0.05 | TRUE | rs3091397 |
| rs11591571 | 4 | chr10:104221149-104418164 | 3.77E-15 | ENSG00000107882 | SUFU | Yes | 0.01 | <0.05 | FALSE | - |
| rs7705526 | 1 | chr5:1253262-1295184 | 6.46E-15 | ENSG00000164362 | TERT | Yes | 0.02 | <0.05 | TRUE | rs2242652;rs6894574 |
| rs12451698;rs3889402 | 10 | chr17:18012020-18317694 | 2.60E-14 | ENSG00000177427 | SMCR7 | Yes | 0.02 | <0.05 | TRUE | rs16960835 |
| rs2304693 | 5 | chr7:45002261-45151646 | 5.11E-11 | ENSG00000136286 | MYO1G | Yes | 0.02 | <0.05 | FALSE | rs6976664 |
| rs1354034 | 1 | chr3:56761446-57113357 | 2.05E-35 | ENSG00000163947 | ARHGEF3 | Yes | 0.02 | <0.05 | FALSE | rs2046823;rs7639049 |
| rs148308452 | 10 | chr12:122277433-122907179 | 3.70E-09 | ENSG00000139719 | VPS33A | Yes | 0.02 | <0.05 | FALSE | - |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000188315 | C3orf62 | Yes | 0.02 | <0.05 | FALSE | rs4955411 |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000105705 | SUGP1 | No | 0.02 | <0.20 | FALSE | - |
| rs1598010 | 2 | chr5:72112139-72386349 | 2.41E-08 | ENSG00000083312 | TNPO1 | No | 0.02 | <0.20 | TRUE | - |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs112997975;rs139111376;rs139898146;rs2375112;rs614997 | 7 | chr1:25549170-25895377 | 1.96E-106 | ENSG00000204178 | TMEM57 | No | 0.02 | <0.20 | FALSE | rs2986161 |
| rs13088724 | 3 | chr3:179040779-179169378 | 7.70E-12 | ENSG00000171109 | MFN1 | No | 0.02 | <0.20 | FALSE | - |
| rs2015599 | 3 | chr12:29302036-29534122 | 6.03E-09 | ENSG00000257176 | FAR2 | No | 0.02 | <0.20 | TRUE | - |
| rs2322718 | 1 | chr8:27168999-27316903 | 1.53E-10 | ENSG00000120899 | PTK2B | No | 0.02 | <0.20 | FALSE | rs2322599;rs2322600 |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000144034 | TPRKB | No | 0.02 | <0.20 | FALSE | - |
| rs4895441 | 1 | chr6:135281516-135424194 | 8.31E-14 | ENSG00000112339 | HBS1L | No | 0.02 | <0.20 | TRUE | rs4472368 |
| rs12451698;rs3889402 | 10 | chr17:18012020-18317694 | 2.60E-14 | ENSG00000176994 | SMCR8 | No | 0.02 | <0.20 | TRUE | - |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000178252 | WDR6 | No | 0.03 | <0.20 | FALSE | rs9846123 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000107862 | GBF1 | No | 0.03 | <0.20 | FALSE | - |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000135617 | PRADC1 | No | 0.03 | <0.20 | FALSE | rs7599223 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000120029 | C10orf76 | No | 0.03 | <0.20 | FALSE | - |
| rs112997975;rs139111376;rs139898146;rs2375112;rs614997 | 7 | chr1:25549170-25895377 | 1.96E-106 | ENSG00000188672 | RHCE | No | 0.03 | <0.20 | FALSE | - |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000198218 | QRICH1 | No | 0.03 | <0.20 | FALSE | rs4974083 |
| rs1613662 | 3 | chr19:55476438-55580914 | 3.90E-09 | ENSG00000022556 | NLRP2 | No | 0.03 | <0.20 | FALSE | rs7253480 |
| rs289713 | 1 | chr16:56995762-57017757 | 4.14E-08 | ENSG00000087237 | CETP | No | 0.03 | <0.20 | TRUE | rs1167742;rs1684575 |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000031823 | RANBP3 | No | 0.03 | <0.20 | FALSE | rs555836;rs1678859 |
| rs13084580 | 6 | chr3:38887260-39234087 | 2.19E-13 | ENSG00000114745 | GORASP1 | No | 0.04 | <0.20 | FALSE | rs11923194 |
| rs139891465;rs35586766;rs56069439;rs7254318 | 8 | chr19:17360838-17453539 | 2.30E-32 | ENSG00000130311 | DDA1 | No | 0.04 | <0.20 | FALSE | - |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000178149 | DALRD3 | No | 0.04 | <0.20 | FALSE | - |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000212123 | PRR22 | No | 0.04 | <0.20 | TRUE | - |
| rs34243225;rs6959832 | 3 | chr7:135046547-135378166 | 2.10E-23 | ENSG00000080802 | CNOT4 | No | 0.04 | <0.20 | TRUE | rs1863004;rs12666242;rs7799891 |
| rs2304693 | 5 | chr7:45002261-45151646 | 5.11E-11 | ENSG00000136280 | CCM2 | No | 0.04 | <0.20 | FALSE | rs2289371;rs1859487;rs11765226 |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000178057 | NDUFAF3 | No | 0.04 | <0.20 | FALSE | rs10865954 |
| rs112997975;rs139111376;rs139898146;rs2375112;rs614997 | 7 | chr1:25549170-25895377 | 1.96E-106 | ENSG00000117614 | SYF2 | No | 0.05 | <0.20 | TRUE | - |
| rs11591571 | 4 | chr10:104221149-104418164 | 3.77E-15 | ENSG00000138111 | TMEM180 | No | 0.06 | <0.20 | FALSE | rs3740416 |
| rs2015599 | 3 | chr12:29302036-29534122 | 6.03E-09 | ENSG00000087502 | ERGIC2 | No | 0.06 | <0.20 | FALSE | rs2278094 |
| rs2015599 | 3 | chr12:29302036-29534122 | 6.03E-09 | ENSG00000087502 | ERGIC2 | No | 0.06 | <0.20 | FALSE | rs2278094 |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000181896 | ZNF101 | No | 0.06 | <0.20 | FALSE | rs247775 |
| rs112997975;rs139111376;rs139898146;rs2375112;rs614997 | 7 | chr1:25549170-25895377 | 1.96E-106 | ENSG00000187010 | RHD | No | 0.07 | <0.20 | TRUE | rs909832 |
| rs139891465;rs35586766;rs56069439;rs7254318 | 8 | chr19:17360838-17453539 | 2.30E-32 | ENSG00000130307 | USHBP1 | No | 0.07 | <0.20 | FALSE | - |
| rs385893 | 1 | chr9:4711155-4742043 | 5.96E-13 | ENSG00000147853 | AK3 | No | 0.08 | <0.20 | TRUE | rs12343429 |
| rs2462124;rs62151973;rs62641680;rs74874677 | 2 | chr2:74119441-74186088 | 4.63E-47 | ENSG00000114956 | DGUOK | No | 0.08 | <0.20 | TRUE | rs4852994;rs6711332;rs13411881 |
| rs148308452 | 10 | chr12:122277433-122907179 | 3.70E-09 | ENSG00000110987 | BCL7A | No | 0.1 | <0.20 | FALSE | - |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000114302 | PRKAR2A | No | 0.1 | <0.20 | FALSE | rs11716614 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs12247015;rs186832534;rs2790203 | 4 | chr10:59951278-60158981 | 1.28E-55 | ENSG00000122873 | CISD1 | No | 0.1 | <0.20 | FALSE | rs1867573;rs1199103 |
| rs10407593;rs11085147;rs9636179 | 7 | chr19:5558178-5720463 | 1.54E-95 | ENSG00000130255 | RPL36 | No | 0.11 | <0.20 | FALSE | rs10406504 |
| rs8067252 | 3 | chr17:29158988-29286340 | 3.58E-08 | ENSG00000184060 | ADAP2 | No | 0.11 | >0.20 | TRUE | - |
| rs2274319 | 1 | chr1:156433519-156470620 | 4.09E-10 | ENSG00000116604 | MEF2D | No | 0.13 | >0.20 | TRUE | rs10908506 |
| rs2977608 | 2 | chr1:761586-789791 | 2.58E-21 | ENSG00000228794 | LINC01128 | No | 0.13 | >0.20 | TRUE | - |
| rs10835540 | 2 | chr11:3875757-4160106 | 5.74E-16 | ENSG00000167323 | STIM1 | No | 0.15 | >0.20 | FALSE | rs7113148 |
| rs445 | 1 | chr7:92234235-92465908 | 8.85E-09 | ENSG00000105810 | CDK6 | No | 0.15 | >0.20 | TRUE | - |
| rs156355;rs4814776;rs669 83532 | 1 | chr20:1875154-1920543 | 2.00E-40 | ENSG00000198053 | SIRPA | No | 0.16 | >0.20 | TRUE | rs6112072 |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000116127 | ALMS1 | No | 0.17 | >0.20 | TRUE | rs6706235 |
| rs13084580 | 6 | chr3:38887260-39234087 | 2.19E-13 | ENSG00000168334 | XIRP1 | No | 0.17 | >0.20 | FALSE | - |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000119915 | ELOVL3 | No | 0.18 | >0.20 | FALSE | - |
| rs114694170 | 2 | chr5:88013975-88762215 | 3.04E-13 | ENSG00000081189 | MEF2C | No | 0.19 | >0.20 | FALSE | rs661311 |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000171124 | FUT3 | No | 0.19 | >0.20 | FALSE | - |
| rs111870993;rs117263028;rs117821325;rs145707459;rs56052501;rs7896518;rs7916282 | 3 | chr10:64893050-65384883 | 9.53E-99 | ENSG00000171988 | JMJD1C | No | 0.19 | >0.20 | TRUE | >0.20 |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000213672 | NCKIPSD | No | 0.23 | >0.20 | TRUE | rs12497850 |
| rs1613662 | 3 | chr19:55476438-55580914 | 3.90E-09 | ENSG00000160439 | RDH13 | No | 0.24 | >0.20 | FALSE | rs3745912 |
| rs148308452 | 10 | chr12:122277433-122907179 | 3.70E-09 | ENSG00000256546 | LOC100506691 | No | 0.24 | >0.20 | FALSE | - |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000068745 | IP6K2 | No | 0.26 | >0.20 | FALSE | - |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000064547 | LPAR2 | No | 0.26 | >0.20 | TRUE | rs880090 |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000156413 | FUT6 | No | 0.26 | >0.20 | FALSE | - |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000087903 | RFX2 | No | 0.27 | >0.20 | FALSE | rs1046391;rs10418205 |
| rs2015599 | 3 | chr12:29302036-29534122 | 6.03E-09 | ENSG00000064763 | FAR2 | No | 0.28 | >0.20 | FALSE | - |
| rs112997975;rs139111376;rs139898146;rs2375112;rs614997 | 7 | chr1:25549170-25895377 | 1.96E-106 | ENSG00000157978 | LDLRAP1 | No | 0.3 | >0.20 | TRUE | rs6688931 |
| rs12247015;rs186832534;rs2790203 | 4 | chr10:59951278-60158981 | 1.28E-55 | ENSG00000151151 | IPMK | No | 0.31 | >0.20 | FALSE | - |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000120049 | KCNIP2 | No | 0.31 | >0.20 | FALSE | - |
| rs12451698;rs3889402 | 10 | chr17:18012020-18317694 | 2.60E-14 | ENSG00000091542 | ALKBH5 | No | 0.31 | >0.20 | FALSE | - |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000171119 | NRTN | No | 0.32 | >0.20 | FALSE | - |
| rs12426673 | 3 | chr12:109460894-109531436 | 4.03E-11 | ENSG00000135093 | USP30 | No | 0.32 | >0.20 | FALSE | - |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000204872 | NAT8B | No | 0.33 | >0.20 | FALSE | - |
| rs13088724 | 3 | chr3:179040779-179169378 | 7.70E-12 | ENSG00000114450 | GNB4 | No | 0.34 | >0.20 | TRUE | rs11714353 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000166171 | DPCD | No | 0.35 | >0.20 | FALSE | rs10883666 |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000105519 | CAPS | No | 0.36 | >0.20 | FALSE | rs6510862;rs8108064 |
| rs6511720 | 1 | chr19:11200038-11244492 | 4.08E-09 | ENSG00000130164 | LDLR | No | 0.36 | >0.20 | TRUE | rs12459603 |
| rs7412 | 2 | chr19:45408956-45422606 | 5.56E-23 | ENSG00000130208 | APOC1 | No | 0.37 | >0.20 | FALSE | - |
| rs11553699;rs9804982 | 4 | chr12:122064455-122232261 | 1.45E-43 | ENSG00000139714 | MORN3 | No | 0.37 | >0.20 | FALSE | - |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000107829 | FBXW4 | No | 0.38 | >0.20 | FALSE | rs17767748 |
| rs12052715 | 1 | chr2:160628362-160761260 | 1.72E-08 | ENSG00000054219 | LY75 | No | 0.42 | >0.20 | FALSE | rs2162500;rs12466631 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000166167 | BTRC | No | 0.43 | >0.20 | FALSE | rs12569599 |
| rs148308452 | 10 | chr12:122277433-122907179 | 3.70E-09 | ENSG00000158104 | HPD | No | 0.45 | >0.20 | TRUE | - |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000213996 | TM6SF2 | No | 0.51 | >0.20 | FALSE | - |
| rs112997975;rs139111376;rs139898146;rs2375112;rs614997 | 7 | chr1:25549170-25895377 | 1.96E-106 | ENSG00000183726 | TMEM50A | No | 0.52 | >0.20 | TRUE | rs1293259 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs4284061 | 1 | chr8:6912831-6914256 | 2.97E-17 | ENSG00000164816 | DEFA5 | No | 0.52 | >0.20 | TRUE | - |
| rs11591571 | 4 | chr10:104221149-104418164 | 3.77E-15 | ENSG00000171206 | TRIM8 | No | 0.53 | >0.20 | TRUE | rs11594073 |
| rs148308452 | 10 | chr12:122277433-122907179 | 3.70E-09 | ENSG00000158023 | WDR66 | No | 0.55 | >0.20 | FALSE | - |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000105717 | PBX4 | No | 0.55 | >0.20 | FALSE | rs12611058 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000107872 | FBXL15 | No | 0.57 | >0.20 | FALSE | - |
| rs148308452 | 10 | chr12:122277433-122907179 | 3.70E-09 | ENSG00000255856 | AC069503.1 | No | 0.57 | >0.20 | FALSE | - |
| rs12148 | 4 | chr22:50946645-50971009 | 1.32E-10 | ENSG00000177989 | ODF3B | No | 0.58 | >0.20 | FALSE | - |
| rs6074896 | 1 | chr20:1544167-1600655 | 1.84E-20 | ENSG00000101307 | SIRPB1 | No | 0.58 | >0.20 | TRUE | - |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000225399 | AC121247.1 | No | 0.59 | >0.20 | FALSE | - |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000198408 | MGEA5 | No | 0.63 | >0.20 | FALSE | rs11191150 |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000135625 | EGR4 | No | 0.63 | >0.20 | FALSE | - |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000144035 | NAT8 | No | 0.64 | >0.20 | FALSE | - |
| rs62494425 | 1 | chr8:9009252-9025646 | 2.07E-08 | ENSG00000253426 | - | No | 0.64 | >0.20 | TRUE | - |
| rs1598010 | 2 | chr5:72112139-72386349 | 2.41E-08 | ENSG00000157107 | FCHO2 | No | 0.65 | >0.20 | FALSE | rs2548332 |
| rs112997975;rs139111376;rs139898146;rs2375112;rs614997 | 7 | chr1:25549170-25895377 | 1.96E-106 | ENSG00000117616 | C1orf63 | No | 0.65 | >0.20 | TRUE | rs592372;rs630931 |
| rs12780745;rs1408343 | 6 | chr10:102672720-102790890 | 6.26E-28 | ENSG00000107816 | LZTS2 | No | 0.65 | >0.20 | TRUE | - |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000163016 | ALMS1P | No | 0.66 | >0.20 | FALSE | - |
| rs111870993;rs117263028;rs117821325;rs145707459;rs56052501;rs7896518;rs7916282 | 3 | chr10:64893050-65384883 | 9.53E-99 | ENSG00000165476 | REEP3 | No | 0.67 | >0.20 | TRUE | rs7076601 |
| rs11064074 | 1 | chr12:6308881-6347427 | 4.87E-20 | ENSG00000010278 | CD9 | No | 0.69 | >0.20 | TRUE | rs11064058 |
| rs114694170 | 2 | chr5:88013975-88762215 | 3.04E-13 | ENSG00000248309 | MEF2C-AS1 | No | 0.71 | >0.20 | TRUE | - |
| rs12780745;rs1408343 | 6 | chr10:102672720-102790890 | 6.26E-28 | ENSG00000186862 | PDZD7 | No | 0.71 | >0.20 | FALSE | - |
| rs2462124;rs62151973;rs62641680;rs74874677 | 2 | chr2:74119441-74186088 | 4.63E-47 | ENSG00000163017 | ACTG2 | No | 0.71 | >0.20 | FALSE | rs6751551 |
| rs148308452 | 10 | chr12:122277433-122907179 | 3.70E-09 | ENSG00000158113 | LRRC43 | No | 0.72 | >0.20 | FALSE | rs7972979 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000059915 | PSD | No | 0.72 | >0.20 | FALSE | - |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000187664 | HAPLN4 | No | 0.74 | >0.20 | FALSE | - |
| rs12451698;rs3889402 | 10 | chr17:18012020-18317694 | 2.60E-14 | ENSG00000131899 | LLGL1 | No | 0.75 | >0.20 | FALSE | rs2290507 |
| rs12780745;rs1408343 | 6 | chr10:102672720-102790890 | 6.26E-28 | ENSG00000095539 | SEMA4G | No | 0.75 | >0.20 | FALSE | - |
| rs55955211 | 4 | chr22:50528308-50656045 | 2.53E-08 | ENSG00000073146 | MOV10L1 | No | 0.76 | >0.20 | FALSE | - |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000144040 | SFXN5 | No | 0.78 | >0.20 | FALSE | rs2115916 |
| rs139891465;rs35586766;rs56069439;rs7254318 | 8 | chr19:17360838-17453539 | 2.30E-32 | ENSG00000074855 | ANO8 | No | 0.78 | >0.20 | FALSE | - |
| rs11697158;rs6105852;rs76599088 | 4 | chr20:17922241-18039832 | 1.41E-25 | ENSG00000125850 | OVOL2 | No | 0.79 | >0.20 | FALSE | rs11299 |
| rs7412 | 2 | chr19:45408956-45422606 | 5.56E-23 | ENSG00000130203 | APOE | No | 0.79 | >0.20 | TRUE | - |
| rs11031389 | 1 | chr11:4208370-4223885 | 3.09E-08 | ENSG00000254480 | LINC02749 | No | 0.79 | >0.20 | TRUE | - |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000173421 | CCDC36 | No | 0.8 | >0.20 | FALSE | rs1134043 |
| rs3766744 | 2 | chr1:11994262-12073571 | 2.91E-17 | ENSG00000083444 | PLOD1 | No | 0.8 | >0.20 | FALSE | rs7550536 |
| rs148308452 | 10 | chr12:122277433-122907179 | 3.70E-09 | ENSG00000176383 | B3GNT4 | No | 0.81 | >0.20 | FALSE | - |
| rs2038480;rs6691993 | 3 | chr1:171810621-172437971 | 1.39E-09 | ENSG00000180999 | C1orf105 | No | 0.85 | >0.20 | FALSE | - |
| rs139891465;rs35586766;rs56069439;rs7254318 | 8 | chr19:17360838-17453539 | 2.30E-32 | ENSG00000127220 | ABHD8 | No | 0.85 | >0.20 | FALSE | rs2288464 |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000249707 | - | No | 0.85 | >0.20 | FALSE | - |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000174898 | TMEM146 | No | 0.86 | >0.20 | TRUE | - |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs112743753 | 2 | chr19:984328-1009731 | 4.79E-08 | ENSG00000116032 | GRIN3B | No | 0.88 | >0.20 | TRUE | rs2285905 |
| rs12426673 | 3 | chr12:109460894-109531436 | 4.03E-11 | ENSG00000256262 | USP30-AS1 | No | 0.89 | >0.20 | TRUE | - |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000163013 | FBXO41 | No | 0.89 | >0.20 | FALSE | - |
| rs2038480;rs6691993 | 3 | chr1:171810621-172437971 | 1.39E-09 | ENSG00000197959 | DNM3 | No | 0.89 | >0.20 | TRUE | - |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000135631 | RAB11FIP5 | No | 0.89 | >0.20 | FALSE | rs13416407 |
| rs11064881 | 3 | chr12:119825792-120315095 | 2.58E-10 | ENSG00000248636 | AC002070.1 | No | 0.9 | >0.20 | TRUE | - |
| rs11064881 | 3 | chr12:119825792-120315095 | 2.58E-10 | ENSG00000122966 | CIT | No | 0.9 | >0.20 | FALSE | - |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000160161 | CILP2 | No | 0.9 | >0.20 | FALSE | - |
| rs1967556 | 1 | chr17:27717943-27871502 | 1.97E-16 | ENSG00000160551 | TAOK1 | No | 0.91 | >0.20 | TRUE | - |
| rs13084580 | 6 | chr3:38887260-39234087 | 2.19E-13 | ENSG00000168026 | TTC21A | No | 0.93 | >0.20 | FALSE | rs11720056 |
| rs117437695;rs3097889 | 12 | chr19:5720688-6110664 | 2.54E-18 | ENSG00000130383 | FUT5 | No | 0.93 | >0.20 | FALSE | - |
| rs10407593;rs11085147;rs9636179 | 7 | chr19:5558178-5720463 | 1.54E-95 | ENSG00000167733 | HSD11B1L | No | 0.93 | >0.20 | FALSE | rs11673407 |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000172037 | LAMB2 | No | 0.94 | >0.20 | FALSE | - |
| rs2038480;rs6691993 | 3 | chr1:171810621-172437971 | 1.39E-09 | ENSG00000135845 | PIGC | No | 0.94 | >0.20 | TRUE | rs13932;rs13932 |
| rs12451698;rs3889402 | 10 | chr17:18012020-18317694 | 2.60E-14 | ENSG00000091536 | MYO15A | No | 0.94 | >0.20 | FALSE | - |
| rs78909033 | 2 | chr2:241505221-241557122 | 1.00E-11 | ENSG00000142330 | CAPN10 | No | 0.95 | >0.20 | FALSE | rs12614493;rs11676358 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000107859 | PITX3 | No | 0.96 | >0.20 | FALSE | rs4919626 |
| rs55955211 | 4 | chr22:50528308-50656045 | 2.53E-08 | ENSG00000073150 | PANX2 | No | 0.96 | >0.20 | FALSE | rs2341367 |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000221883 | C3orf71 | No | 0.96 | >0.20 | FALSE | - |
| rs116614177 | 15 | chr2:73169165-74007284 | 2.17E-09 | ENSG00000246432 | - | No | 0.97 | >0.20 | FALSE | - |
| rs13381785;rs77261872 | 1 | chr18:42260138-42648475 | 2.28E-16 | ENSG00000152217 | SETBP1 | No | 0.97 | >0.20 | TRUE | rs1036929;rs4890486 |
| rs2304693 | 5 | chr7:45002261-45151646 | 5.11E-11 | ENSG00000136274 | NACAD | No | 0.97 | >0.20 | FALSE | - |
| rs12451698;rs3889402 | 10 | chr17:18012020-18317694 | 2.60E-14 | ENSG00000214860 | EVPLL | No | 0.98 | >0.20 | FALSE | - |
| rs2304128 | 12 | chr19:19366456-19791761 | 9.16E-09 | ENSG00000178093 | TSSK6 | No | 0.98 | >0.20 | FALSE | - |
| rs13084580 | 6 | chr3:38887260-39234087 | 2.19E-13 | ENSG00000168356 | SCN11A | No | 0.98 | >0.20 | FALSE | - |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000226009 | KCNIP2-AS1 | No | 0.98 | >0.20 | FALSE | - |
| rs4841132 | 1 | chr8:9106927-9271224 | 7.46E-09 | ENSG00000254235 | - | No | 0.98 | >0.20 | FALSE | - |
| rs10407593;rs11085147;rs9636179 | 7 | chr19:5558178-5720463 | 1.54E-95 | ENSG00000223573 | PLAC2 | No | 0.99 | >0.20 | TRUE | - |
| rs72698722 | 1 | chr14:101192042-101201539 | 1.25E-12 | ENSG00000185559 | DLK1 | No | 0.99 | >0.20 | TRUE | rs12882815;rs10138818 |
| rs12451698;rs3889402 | 10 | chr17:18012020-18317694 | 2.60E-14 | ENSG00000220161 | LINC02076 | No | 0.99 | >0.20 | FALSE | - |
| rs4727801;rs6972244 | 3 | chr7:113726382-114766368 | 2.77E-25 | ENSG00000225535 | LINC01393 | No | 0.99 | >0.20 | TRUE | - |
| rs4727801;rs6972244 | 3 | chr7:113726382-114766368 | 2.77E-25 | ENSG00000128573 | FOXP2 | No | 0.99 | >0.20 | FALSE | rs1058335;rs6945523 |
| rs5012419 | 1 | chr12:19282648-19529334 | 2.66E-29 | ENSG00000052126 | PLEKHA5 | No | 1 | >0.20 | TRUE | rs12372773 |
| rs59791256 | 21 | chr10:103113820-104192418 | 3.60E-10 | ENSG00000107831 | FGF8 | No | 1 | >0.20 | FALSE | rs3127245 |
| rs6792510 | 20 | chr3:48701364-49378145 | 1.44E-08 | ENSG00000178467 | P4HTM | No | 1 | >0.20 | FALSE | - |
| rs2018800;rs465339 | 1 | chr19:5455426-5456867 | 7.81E-09 | ENSG00000105428 | ZNRF4 | No | 1 | >0.20 | TRUE | - |
| rs1569419 | 1 | chr1:2985732-3355185 | 6.87E-14 | ENSG00000142611 | PRDM16 | No | 1 | >0.20 | TRUE | rs10492937 |

**S2. Table 5**. MitoCarta3 annotations for DEPICT and GeneMANIA-prioritized genes

| Gene Overview | | | MitoCarta3.0 Annotations | | | | | |
|---|---|---|---|---|---|---|---|---|
| Gene Name | Analysis Identified | Human MitoCarta3 Status | MitoCarta3.0 Gene Name | MitoCarta3.0_Sub MitoLocalization | MitoCarta3.0 MitoPathways | UniProt | PGC Induction Score | Tissues Expressed |
| BIK | DEPICT | Yes | BIK | MOM | Mitochondrial dynamics and surveillance > Apoptosis | Q13323 | NA | NA |
| CLPX | GeneMANIA | Yes | CLPX | Matrix | Protein import, sorting and homeostasis > Protein homeostasis > Proteases | O76031 | 0.68 | all 14 |
| EARS2 | GeneMANIA | Yes | EARS2 | Matrix | Mitochondrial central dogma > Translation > mt-tRNA synthetases | Q5JPH6 | NA | adipose, largeintestine, placenta |
| GARS1 | GeneMANIA | Yes | GARS1 | Matrix | Mitochondrial central dogma > Translation > mt-tRNA synthetases | P41250 | 1.62 | cerebrum, cerebellum, brainstem, spinalcord, liver, skeletalmuscle, adipose, smallintestine, largeintestine, stomach, placenta, testis |
| GTPBP3 | DEPICT | Yes | GTPBP3 | Matrix | Mitochondrial central dogma > mtRNA metabolism > mt-tRNA modifications \| Metabolism > Vitamin metabolism > Folate and 1-C metabolism | Q969Y2 | NA | largeintestine |
| LONP1 | DEPICT | Yes | LONP1 | Matrix | Protein import, sorting and homeostasis > Protein homeostasis > Proteases | P36776 | 1.81 | all 14 |
| PIM1 | GeneMANIA | Yes | LONP1 | Matrix | Protein import, sorting and homeostasis > Protein homeostasis > Proteases | P36776 | 1.81 | all 14 |
| MCAT | DEPICT | Yes | MCAT | Matrix | Metabolism > Lipid metabolism > Type II fatty acid synthesis | Q8IVS2 | NA | spinalcord, kidney, liver, skeletalmuscle, adipose, smallintestine, largeintestine, stomach, placenta, testis |
| MFN2 | DEPICT | Yes | MFN2 | MOM | Mitochondrial dynamics and surveillance > Fusion \| Mitochondrial dynamics and surveillance > Organelle contact sites | O95140 | 1.83 | cerebrum, cerebellum, brainstem, spinalcord, liver, adipose, largeintestine, stomach, placenta, testis |
| C20orf72; MGME1 | DEPICT | Yes | MGME1 | Matrix | Mitochondrial central dogma > mtDNA maintenance > mtDNA replication \| Mitochondrial central dogma > mtDNA maintenance > mtDNA stability and decay | Q9BQP7 | NA | largeintestine, placenta |
| C19orf70; MICOS13 | DEPICT | Yes | MICOS13 | MIM | Mitochondrial dynamics and surveillance > Cristae formation > MICOS complex | Q5XKP0 | 0.72 | all 14 |
| SMCR7 | DEPICT | Yes | MIEF2 | MOM | Mitochondrial dynamics and surveillance > Fission | Q96C03 | NA | NA |
| MRPL34 | DEPICT | Yes | MRPL34 | Matrix | Mitochondrial central dogma > Translation > Mitochondrial ribosome | Q9BQ48 | 1.75 | cerebrum, cerebellum, brainstem, spinalcord, kidney, liver, heart, adipose, smallintestine, largeintestine, stomach, placenta, testis |
| MRPL43 | DEPICT | Yes | MRPL43 | Matrix | Mitochondrial central dogma > Translation > Mitochondrial ribosome | Q8N983 | 1.57 | all 14 |
| MRPS35 | DEPICT | Yes | MRPS35 | Matrix | Mitochondrial central dogma > Translation > Mitochondrial ribosome | P82673 | 1.81 | all 14 |
| NDUFA11 | DEPICT | Yes | NDUFA11 | MIM | OXPHOS > Complex I > CI subunits \| OXPHOS > OXPHOS subunits | Q86Y39 | 1.8 | all 14 |
| NDUFA7 | GeneMANIA | Yes | NDUFA7 | MIM | OXPHOS > Complex I > CI subunits \| OXPHOS > OXPHOS subunits | O95182 | NA | all 14 |
| NOA1 | GeneMANIA | Yes | NOA1 | MIM | Mitochondrial central dogma > mtRNA metabolism > mtRNA granules \| Mitochondrial central dogma > Translation > Mitochondrial ribosome assembly | Q8NC60 | NA | adipose, largeintestine |
| OXA1L | GeneMANIA | Yes | OXA1L | MIM | Protein import, sorting and homeostasis > Protein import and sorting > OXA \| Mitochondrial central dogma > Translation | Q15070 | NA | cerebrum, brainstem, spinalcord, kidney, heart, adipose, smallintestine, largeintestine, placenta, testis |
| POLG2 | DEPICT | Yes | POLG2 | Matrix | Mitochondrial central dogma > mtDNA maintenance > mtDNA replication \| Mitochondrial central dogma > mtDNA maintenance > mtDNA nucleoid | Q9UHN1 | NA | NA |
| SCO2 | DEPICT | Yes | SCO2 | MIM | OXPHOS > Complex IV > CIV assembly factors \| Metabolism > Metals and cofactors > Copper metabolism \| OXPHOS > OXPHOS assembly factors | O43819 | NA | all 14 |
| SELO | DEPICT | Yes | SELENOO | Matrix | Metabolism > Detoxification > Selenoproteins | Q9BVL4 | NA | spinalcord, liver, skeletalmuscle, adipose, smallintestine, largeintestine, stomach, placenta, testis |
| SLC25A20 | DEPICT | Yes | SLC25A20 | MIM | Metabolism > Lipid metabolism > Fatty acid oxidation \| Metabolism > Metals and cofactors > Carnitine synthesis and transport \| Metabolism > Metals and cofactors > Carnitine shuttle \| Small molecule transport > SLC25A family | O43772 | 1.62 | all 14 |
| TBRG4 | DEPICT | Yes | TBRG4 | Matrix | Mitochondrial central dogma > mtRNA metabolism > Polycistronic mtRNA processing \| Mitochondrial central dogma > mtRNA metabolism > mtRNA stability and decay | Q969Z0 | NA | all 14 |
| C17orf42; TEFM | DEPICT | Yes | TEFM | Matrix | Mitochondrial central dogma > mtRNA metabolism > Transcription | Q96QE5 | 1.14 | liver, heart, skeletalmuscle, adipose, largeintestine, stomach, placenta, testis |
| TFAM | DEPICT | Yes | TFAM | Matrix | Mitochondrial central dogma > mtDNA maintenance > mtDNA replication \| Mitochondrial central dogma > mtDNA maintenance > mtDNA nucleoid \| Mitochondrial central dogma > mtRNA metabolism > Transcription | Q00059 | 1.62 | all 14 |
| TOP3A | DEPICT | Yes | TOP3A | Matrix | Mitochondrial central dogma > mtDNA maintenance > mtDNA replication | Q13472 | NA | testis |
| C10orf2; TWNK | DEPICT | Yes | TWNK | Matrix | Mitochondrial central dogma > mtDNA maintenance > mtDNA replication \| Mitochondrial central dogma > mtDNA maintenance > mtDNA nucleoid | Q96RR1 | 1.99 | NA |
| ACTR1A | DEPICT | No | ACTR1A | NA | NA | P61163 | NA | cerebrum, cerebellum, brainstem, spinalcord, kidney, liver, skeletalmuscle, adipose, smallintestine, largeintestine, stomach, placenta, testis |
| ALKBH2 | DEPICT | No | ALKBH2 | NA | NA | Q6N538 | NA | NA |
| ANKLE1 | DEPICT | No | ANKLE1 | NA | NA | Q8NAG6 | NA | NA |
| ARIH2 | DEPICT | No | ARIH2 | NA | NA | O95376 | NA | NA |
| ATAD5 | DEPICT | No | ATAD5 | NA | NA | Q96QE3 | NA | largeintestine, testis |
| ATP13A1 | DEPICT | No | ATP13A1 | NA | NA | Q9HD20 | NA | testis |
| BABAM1 | DEPICT | No | BABAM1 | NA | NA | Q9NWV8 | NA | NA |
| BANK1 | DEPICT | No | BANK1 | NA | NA | Q8NDB2 | NA | NA |
| C3orf62 | DEPICT | No | C3orf62 | NA | NA | Q6ZUJ4 | NA | NA |
| C7orf74 | DEPICT | No | CCDC71L | NA | NA | Q8N9Z2 | NA | NA |
| CCT7 | DEPICT | No | CCT7 | NA | NA | Q99832 | -0.3 | cerebrum, spinalcord, adipose, smallintestine, placenta, testis |
| CD226 | DEPICT | No | CD226 | NA | NA | Q15762 | NA | NA |
| CD36 | DEPICT | No | CD36 | NA | NA | P16671 | NA | adipose, smallintestine |
| CLIP1 | DEPICT | No | CLIP1 | NA | NA | P30622 | NA | skeletalmuscle, largeintestine |
| CPNE3 | GeneMANIA | No | CPNE3 | NA | NA | O75131 | NA | NA |
| HN1L | GeneMANIA | No | CRAMP1 | NA | NA | Q96RY5 | NA | largeintestine |
| CSRNP1 | DEPICT | No | CSRNP1 | NA | NA | Q96S65 | 1.39 | NA |
| CUEDC2 | DEPICT | No | CUEDC2 | NA | NA | Q9H467 | NA | NA |
| DRG2 | DEPICT | No | DRG2 | NA | NA | P55039 | 1.28 | NA |
| DUS3L | DEPICT | No | DUS3L | NA | NA | Q96G46 | NA | NA |
| DUSP11 | DEPICT | No | DUSP11 | NA | NA | O75319 | NA | NA |
| ARHGEF3 | DEPICT | No | ECT2 | NA | NA | Q9H8V3 | NA | NA |
| FLII | DEPICT | No | FLII | NA | NA | Q13045 | -0.23 | NA |
| GATAD2A | DEPICT | No | GATAD2A | NA | NA | Q86YP4 | NA | NA |
| GMIP | DEPICT | No | GMIP | NA | NA | Q9P107 | NA | adipose |
| GP6 | DEPICT | No | GP6 | NA | NA | Q9HCN6 | NA | NA |
| HPS5 | GeneMANIA | No | HPS5 | NA | NA | Q9UPZ3 | NA | NA |
| HPS6 | DEPICT | No | HPS6 | NA | NA | Q86YV9 | NA | NA |
| IMPDH2 | DEPICT | No | IMPDH2 | NA | NA | P12268 | NA | NA |
| ITSN1 | GeneMANIA | No | ITSN1 | NA | NA | Q15811 | NA | cerebellum, brainstem |
| KLHDC8B | DEPICT | No | KLHDC8B | NA | NA | Q8IXV7 | NA | NA |
| LDB1 | DEPICT | No | LDB1 | NA | NA | Q86U70 | NA | NA |
| MARK4 | GeneMANIA | No | MARK4 | NA | NA | Q96L34 | NA | NA |
| MAU2 | DEPICT | No | MAU2 | NA | NA | Q9Y6X3 | NA | NA |
| MLXIP | DEPICT | No | MLXIP | NA | NA | Q9HAP2 | NA | NA |
| MYO1G | DEPICT | No | MYO1G | NA | NA | B0I1T2 | NA | smallintestine |
| NCAPH2 | DEPICT | No | NCAPH2 | NA | NA | Q6IBW4 | -0.56 | NA |
| NFKB2 | DEPICT | No | NFKB2 | NA | NA | Q00653 | NA | NA |
| NOLC1 | DEPICT | No | NOLC1 | NA | NA | Q14978 | NA | NA |
| NPM3 | DEPICT | No | NPM3 | NA | NA | O75607 | NA | NA |
| NRBF2 | DEPICT | No | NRBF2 | NA | NA | Q96F24 | NA | NA |
| NUP205 | DEPICT | No | NUP205 | NA | NA | Q92621 | NA | testis |
| ORAI1 | DEPICT | No | ORAI1 | NA | NA | Q96D31 | NA | NA |
| PAK4 | GeneMANIA | No | PAK4 | NA | NA | O96013 | NA | NA |
| PIN1 | GeneMANIA | No | PIN1 | NA | NA | Q13526 | NA | brainstem, spinalcord |
| PNP | DEPICT | No | PNP | NA | NA | P00491 | NA | smallintestine, placenta |
| POLL | DEPICT | No | POLL | NA | NA | Q9UGP5 | NA | NA |
| PPRC1 | DEPICT | No | PPRC1 | NA | NA | Q5VV67 | -0.48 | NA |
| PRKAB1 | DEPICT | No | PRKAB1 | NA | NA | Q9Y478 | NA | NA |
| PUM3 | GeneMANIA | No | PUM3 | NA | NA | Q15397 | NA | NA |
| QARS | DEPICT | No | QARS1 | NA | NA | P47897 | -0.39 | NA |
| RHOF | DEPICT | No | RHOF | NA | NA | Q9HBH0 | NA | NA |
| RNPEPL1 | DEPICT | No | RNPEPL1 | NA | NA | Q9HAU8 | NA | NA |
| RRM1 | DEPICT | No | RRM1 | NA | NA | P23921 | 0.82 | NA |
| SAFB | DEPICT | No | SAFB | NA | NA | Q15424 | NA | NA |
| SAFB2 | DEPICT | No | SAFB2 | NA | NA | Q14151 | NA | NA |
| SHMT1 | DEPICT | No | SHMT1 | NA | NA | P34896 | 2.69 | NA |
| FAM178A | DEPICT | No | SLF2 | NA | NA | Q8IX21 | NA | NA |
| SMYD5 | DEPICT | No | SMYD5 | NA | NA | Q6GMV2 | NA | NA |
| SNX5 | DEPICT | No | SNX5 | NA | NA | Q9Y5X3 | -0.41 | NA |
| SPCS2 | GeneMANIA | No | SPCS2 | NA | NA | Q15005 | -0.42 | liver, placenta |
| C7orf73 | DEPICT | No | STMP1 | NA | NA | E0CX11 | NA | NA |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SUFU | DEPICT | No | SUFU | NA | NA | Q9UMX1 | NA | NA |
| TERT | DEPICT | No | TERT | NA | NA | Q14746 | NA | NA |
| TMEM120B | DEPICT | No | TMEM120B | NA | NA | A0PK00 | NA | NA |
| TP53BP2 | GeneMANIA | No | TP53BP2 | NA | NA | Q13625 | NA | NA |
| TRABD | DEPICT | No | TRABD | NA | NA | Q9H4I3 | 0.5 | kidney, largeintestine, placenta, testis |
| TRIM58 | DEPICT | No | TRIM58 | NA | NA | Q8NG06 | NA | NA |
| TYMP | DEPICT | No | TYMP | NA | NA | P19971 | NA | NA |
| UBE2D1 | DEPICT | No | UBE2D1 | NA | NA | P51668 | NA | NA |
| USP19 | DEPICT | No | USP19 | NA | NA | O94966 | NA | NA |
| USP4 | DEPICT | No | USP4 | NA | NA | Q13107 | 0.41 | NA |
| VPS11 | GeneMANIA | No | VPS11 | NA | NA | Q9H270 | NA | NA |
| VPS16 | GeneMANIA | No | VPS16 | NA | NA | Q9H269 | NA | NA |
| VPS18 | GeneMANIA | No | VPS18 | NA | NA | Q9P253 | NA | cerebrum, cerebellum, brainstem, adipose, placenta, testis |
| VPS33A | DEPICT | No | VPS33A | NA | NA | Q96AX1 | NA | NA |
| WDR18 | DEPICT | No | WDR18 | NA | NA | Q9BV38 | NA | NA |
| WDR48 | DEPICT | No | WDR48 | NA | NA | Q8TAF3 | NA | NA |
| ZNF639 | DEPICT | No | ZNF639 | NA | NA | Q9UID6 | 0.87 | NA |

**S2. Table 6.** Rare variant exome-wide association testing for mtDNA-CN loci. Nominally significant (P<0.01) results are shown.

| Index | Gene | Carrier Count | Non-carrier Count | beta | se | 95% LCI | 95% UCI | P-value |
|---|---|---|---|---|---|---|---|---|
| 14073 | SAMHD1 | 1112 | 146628 | 0.234084 | 0.02606 | 0.183017 | 0.285151 | 2.63E-19 |
| 16251 | TFAM | 148 | 147592 | -0.33052 | 0.07186 | -0.47136 | -0.18969 | 4.23E-06 |
| 17910 | WDR59 | 1253 | 146487 | -0.10931 | 0.0248 | -0.15791 | -0.06071 | 1.04E-05 |
| 2834 | CELF5 | 286 | 147454 | 0.224476 | 0.05172 | 0.123104 | 0.325848 | 1.42E-05 |
| 12513 | PPP1R35 | 304 | 147436 | 0.192313 | 0.0459 | 0.102341 | 0.282284 | 2.80E-05 |
| 538 | ALDH16A1 | 1925 | 145815 | 0.078944 | 0.01905 | 0.041614 | 0.116273 | 3.40E-05 |
| 15623 | STAG2 | 154 | 147586 | -0.29098 | 0.07049 | -0.42913 | -0.15282 | 3.66E-05 |
| 11997 | PIGT | 744 | 146996 | -0.13189 | 0.03205 | -0.19471 | -0.06906 | 3.88E-05 |
| 14753 | SLC25A37 | 302 | 147438 | 0.20316 | 0.05034 | 0.104486 | 0.301834 | 5.45E-05 |
| 1211 | ATG9A | 590 | 147150 | 0.141701 | 0.03595 | 0.071231 | 0.212172 | 8.11E-05 |
| 1709 | BTNL2 | 249 | 147491 | 0.218173 | 0.05543 | 0.109531 | 0.326814 | 8.29E-05 |
| 15945 | TAFA2 | 93 | 147647 | 0.34432 | 0.09066 | 0.166629 | 0.522011 | 0.000146 |
| 297 | ADAR | 748 | 146992 | 0.118711 | 0.03147 | 0.057028 | 0.180394 | 0.000162 |
| 15735 | STX1A | 134 | 147606 | 0.275124 | 0.07552 | 0.127099 | 0.423149 | 0.00027 |
| 7184 | HRH4 | 223 | 147517 | 0.202505 | 0.05564 | 0.093458 | 0.311553 | 0.000273 |
| 3267 | CLUH | 2468 | 145272 | -0.06281 | 0.01732 | -0.09676 | -0.02886 | 0.000287 |
| 16182 | TEK | 711 | 147029 | -0.11051 | 0.03285 | -0.17489 | -0.04614 | 0.000767 |
| 903 | AQP9 | 512 | 147228 | 0.130133 | 0.03869 | 0.054297 | 0.205969 | 0.00077 |
| 994 | ARHGEF6 | 266 | 147474 | 0.178188 | 0.05366 | 0.073015 | 0.28336 | 0.000898 |
| 9877 | MT1M | 13 | 147727 | -0.80468 | 0.24236 | -1.2797 | -0.32967 | 0.0009 |
| 3659 | CRYM | 224 | 147516 | -0.19254 | 0.05805 | -0.30631 | -0.07876 | 0.00091 |
| 5900 | FXYD4 | 95 | 147645 | -0.29724 | 0.08968 | -0.47301 | -0.12148 | 0.000918 |
| 4314 | DLGAP3 | 397 | 147343 | -0.1434 | 0.0436 | -0.22885 | -0.05796 | 0.001004 |
| 13664 | RNF126 | 949 | 146791 | 0.093211 | 0.02837 | 0.037615 | 0.148807 | 0.001016 |
| 8799 | LPXN | 685 | 147055 | -0.10371 | 0.03164 | -0.16573 | -0.04169 | 0.001047 |
| 13164 | RABL2B | 321 | 147419 | 0.158845 | 0.04883 | 0.063146 | 0.254544 | 0.001141 |
| 15978 | TAS1R1 | 1947 | 145793 | -0.06317 | 0.01966 | -0.1017 | -0.02465 | 0.001309 |
| 14665 | SLC16A9 | 388 | 147352 | 0.139698 | 0.04375 | 0.053954 | 0.225443 | 0.001407 |
| 7496 | IL1RAP | 167 | 147573 | 0.214074 | 0.06765 | 0.081471 | 0.346677 | 0.001555 |
| 6423 | GPATCH2 | 326 | 147414 | -0.15327 | 0.04845 | -0.24823 | -0.05831 | 0.001559 |
| 7817 | JOSD1 | 57 | 147683 | -0.36576 | 0.11575 | -0.59263 | -0.13888 | 0.001579 |
| 6994 | HIST1H4E | 324 | 147416 | -0.15336 | 0.04863 | -0.24867 | -0.05806 | 0.001611 |
| 7628 | INTS8 | 275 | 147465 | -0.16592 | 0.05274 | -0.26929 | -0.06254 | 0.001656 |
| 15794 | SUPT4H1 | 88 | 147652 | -0.29177 | 0.09318 | -0.47441 | -0.10913 | 0.001742 |
| 8900 | LRRC74A | 434 | 147306 | 0.129644 | 0.04159 | 0.048136 | 0.211152 | 0.001824 |
| 14290 | SEMA4B | 1744 | 145996 | -0.06511 | 0.0209 | -0.10607 | -0.02414 | 0.001839 |
| 5703 | FLT3 | 811 | 146929 | 0.095601 | 0.03073 | 0.035377 | 0.155826 | 0.001863 |
| 13432 | REPIN1 | 579 | 147161 | 0.11261 | 0.03629 | 0.041475 | 0.183746 | 0.001918 |
| 15576 | SSTR3 | 557 | 147183 | 0.11446 | 0.03699 | 0.041953 | 0.186968 | 0.001975 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 10904 | ODF3L1 | 313 | 147427 | -0.15226 | 0.04946 | -0.24921 | -0.05532 | 0.002081 |
| 12376 | POMP | 55 | 147685 | 0.360846 | 0.11784 | 0.129878 | 0.591814 | 0.002198 |
| 15485 | SPTLC2 | 458 | 147282 | 0.119405 | 0.03902 | 0.042924 | 0.195886 | 0.002214 |
| 9801 | MRPS23 | 110 | 147630 | 0.252238 | 0.08334 | 0.088888 | 0.415587 | 0.002474 |
| 10018 | MYH2 | 3135 | 144605 | -0.04677 | 0.01546 | -0.07707 | -0.01646 | 0.002489 |
| 10787 | NUDCD1 | 245 | 147495 | 0.16675 | 0.05519 | 0.058571 | 0.274928 | 0.002518 |
| 9042 | MAB21L4 | 325 | 147415 | -0.14465 | 0.04794 | -0.23861 | -0.05068 | 0.002553 |
| 11209 | OR5AK2 | 91 | 147649 | -0.27598 | 0.09164 | -0.45559 | -0.09637 | 0.002598 |
| 2257 | CASP1 | 167 | 147573 | 0.202762 | 0.06765 | 0.07016 | 0.335364 | 0.002727 |
| 868 | APOC2 | 139 | 147601 | 0.222172 | 0.07417 | 0.076799 | 0.367544 | 0.002741 |
| 7353 | IFI35 | 353 | 147387 | 0.138124 | 0.04637 | 0.047234 | 0.229013 | 0.002896 |
| 14389 | SETD7 | 346 | 147394 | 0.140176 | 0.0471 | 0.047863 | 0.232488 | 0.002919 |
| 16334 | THOC2 | 146 | 147594 | 0.214448 | 0.07238 | 0.07259 | 0.356307 | 0.003048 |
| 18564 | ZNF519 | 390 | 147350 | 0.130834 | 0.04431 | 0.043994 | 0.217674 | 0.003148 |
| 10510 | NLRP8 | 456 | 147284 | -0.12028 | 0.04086 | -0.20036 | -0.0402 | 0.00324 |
| 14037 | S1PR2 | 265 | 147475 | 0.158156 | 0.05373 | 0.052838 | 0.263473 | 0.003247 |
| 18542 | ZNF490 | 53 | 147687 | -0.31345 | 0.10677 | -0.52272 | -0.10418 | 0.003328 |
| 11244 | OR5M9 | 46 | 147694 | 0.377561 | 0.12888 | 0.124963 | 0.63016 | 0.003394 |
| 16718 | TMEM72 | 115 | 147625 | -0.23882 | 0.08153 | -0.39861 | -0.07903 | 0.003396 |
| 5543 | FCGR2B | 33 | 147707 | -0.44538 | 0.15212 | -0.74354 | -0.14721 | 0.003415 |
| 8316 | KRT80 | 1155 | 146585 | 0.074957 | 0.02561 | 0.024754 | 0.12516 | 0.003429 |
| 7621 | INTS2 | 921 | 146819 | -0.08379 | 0.0287 | -0.14004 | -0.02754 | 0.003503 |
| 15832 | SYCP1 | 331 | 147409 | -0.14033 | 0.04808 | -0.23458 | -0.04609 | 0.003517 |
| 4014 | DCAF17 | 468 | 147272 | 0.1146 | 0.03938 | 0.037407 | 0.191793 | 0.003617 |
| 1569 | BLOC1S6 | 35 | 147705 | -0.42921 | 0.14772 | -0.71873 | -0.13969 | 0.003665 |
| 6632 | GSE1 | 3397 | 144343 | 0.043108 | 0.01485 | 0.014007 | 0.072209 | 0.003692 |
| 17311 | TUBB1 | 1243 | 146497 | 0.071977 | 0.02482 | 0.023334 | 0.120621 | 0.00373 |
| 6837 | HDDC2 | 539 | 147201 | 0.108939 | 0.03761 | 0.035226 | 0.182653 | 0.003772 |
| 3694 | CSNK2A2 | 61 | 147679 | -0.32331 | 0.11191 | -0.54266 | -0.10397 | 0.003864 |
| 18373 | ZNF208 | 428 | 147312 | 0.122058 | 0.0423 | 0.039155 | 0.204961 | 0.003906 |
| 18628 | ZNF595 | 2169 | 145571 | -0.03191 | 0.01106 | -0.05358 | -0.01023 | 0.003919 |
| 9028 | LYZL4 | 127 | 147613 | 0.223582 | 0.07757 | 0.071554 | 0.375611 | 0.003946 |
| 13071 | QSOX2 | 1053 | 146687 | 0.076902 | 0.02669 | 0.024599 | 0.129205 | 0.003955 |
| 14227 | SDR16C5 | 617 | 147123 | 0.101223 | 0.03516 | 0.032306 | 0.17014 | 0.003993 |
| 16556 | TMEM156 | 71 | 147669 | -0.29728 | 0.10373 | -0.50059 | -0.09398 | 0.004157 |
| 11848 | PFDN4 | 42 | 147698 | -0.38635 | 0.13485 | -0.65066 | -0.12205 | 0.00417 |
| 18227 | ZEB1 | 437 | 147303 | -0.11878 | 0.04147 | -0.20007 | -0.03749 | 0.004185 |
| 13319 | RBM15B | 914 | 146826 | -0.08156 | 0.02853 | -0.13748 | -0.02565 | 0.00425 |
| 976 | ARHGEF15 | 1038 | 146702 | 0.077559 | 0.02714 | 0.024357 | 0.130761 | 0.004273 |
| 4616 | DYM | 1442 | 146298 | 0.06545 | 0.02291 | 0.020549 | 0.110351 | 0.004277 |
| 3455 | COPRS | 147 | 147593 | 0.205838 | 0.07211 | 0.064506 | 0.347169 | 0.00431 |
| 16684 | TMEM38B | 428 | 147312 | -0.12061 | 0.0423 | -0.20352 | -0.0377 | 0.004356 |
| 7552 | IMP3 | 188 | 147552 | 0.173682 | 0.06093 | 0.054266 | 0.293099 | 0.004363 |
| 8035 | KIAA0754 | 976 | 146764 | -0.07945 | 0.02789 | -0.13412 | -0.02478 | 0.004393 |
| 5500 | FBXO28 | 56 | 147684 | 0.332062 | 0.11678 | 0.103168 | 0.560956 | 0.004464 |
| 17348 | TWNK | 1328 | 146412 | -0.06734 | 0.0237 | -0.11378 | -0.02089 | 0.004489 |
| 12233 | PLSCR5 | 246 | 147494 | -0.1569 | 0.05543 | -0.26554 | -0.04827 | 0.004642 |
| 8752 | LOC11448 | 259 | 147481 | -0.15334 | 0.05435 | -0.25986 | -0.04682 | 0.00478 |
| 8388 | KRTAP5-4 | 17 | 147723 | 0.597892 | 0.21199 | 0.182388 | 1.013395 | 0.004798 |
| 8772 | LONRF2 | 801 | 146939 | 0.086987 | 0.03085 | 0.026528 | 0.147446 | 0.004803 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5536 | FCER1A | 44 | 147696 | -0.371 | 0.13173 | -0.6292 | -0.11281 | 0.004858 |
| 6194 | GIMAP6 | 121 | 147619 | -0.22372 | 0.07948 | -0.37949 | -0.06794 | 0.00488 |
| 7991 | KDM1A | 385 | 147355 | 0.123532 | 0.04392 | 0.037458 | 0.209605 | 0.004909 |
| 10971 | OR10A6 | 177 | 147563 | 0.183088 | 0.06519 | 0.055321 | 0.310855 | 0.004976 |
| 4271 | DIAPH2 | 385 | 147355 | -0.12433 | 0.0443 | -0.21116 | -0.03751 | 0.005006 |
| 8353 | KRTAP19-6 | 30 | 147710 | 0.447676 | 0.15955 | 0.134953 | 0.7604 | 0.005019 |
| 13429 | REN | 510 | 147230 | 0.108681 | 0.03876 | 0.032704 | 0.184658 | 0.005053 |
| 18383 | ZNF222 | 153 | 147587 | 0.197948 | 0.07071 | 0.059366 | 0.33653 | 0.005117 |
| 18807 | ZNF85 | 105 | 147635 | -0.23876 | 0.08531 | -0.40597 | -0.07155 | 0.005131 |
| 13433 | REPS1 | 371 | 147369 | 0.127104 | 0.04542 | 0.038083 | 0.216124 | 0.005135 |
| 17387 | U2AF2 | 167 | 147573 | -0.18923 | 0.06765 | -0.32182 | -0.05664 | 0.005156 |
| 7808 | JHY | 482 | 147258 | -0.11144 | 0.03987 | -0.18958 | -0.0333 | 0.005186 |
| 3833 | CXCL11 | 93 | 147647 | 0.253202 | 0.09064 | 0.075556 | 0.430848 | 0.005213 |
| 9228 | MARVELD | 184 | 147556 | 0.179769 | 0.06445 | 0.053439 | 0.306099 | 0.005286 |
| 11823 | PES1 | 955 | 146785 | 0.078432 | 0.02815 | 0.023265 | 0.133598 | 0.005328 |
| 15639 | STARD7 | 459 | 147281 | 0.113291 | 0.04072 | 0.033483 | 0.193099 | 0.005398 |
| 18066 | YBX2 | 280 | 147460 | 0.144758 | 0.05227 | 0.042312 | 0.247204 | 0.005615 |
| 718 | ANKRD34E | 181 | 147559 | 0.180134 | 0.06506 | 0.052626 | 0.307641 | 0.005624 |
| 16413 | TKTL2 | 1011 | 146729 | -0.07577 | 0.02737 | -0.12943 | -0.02212 | 0.005638 |
| 11100 | OR2T5 | 24 | 147716 | -0.46535 | 0.16816 | -0.79494 | -0.13577 | 0.005651 |
| 1155 | ASPDH | 469 | 147271 | -0.11038 | 0.03995 | -0.18867 | -0.03208 | 0.005726 |
| 6584 | GRIK2 | 547 | 147193 | 0.103033 | 0.03733 | 0.029858 | 0.176208 | 0.005785 |
| 4397 | DNAJC13 | 1282 | 146458 | -0.06617 | 0.02399 | -0.11318 | -0.01916 | 0.0058 |
| 6106 | GCSAML | 44 | 147696 | -0.36322 | 0.13176 | -0.62146 | -0.10498 | 0.005838 |
| 3634 | CRTC1 | 306 | 147434 | 0.137753 | 0.05001 | 0.039725 | 0.235781 | 0.005883 |
| 6115 | GDF11 | 1141 | 146599 | 0.071411 | 0.02593 | 0.020585 | 0.122238 | 0.005891 |
| 16925 | TPM1 | 443 | 147297 | 0.112257 | 0.04081 | 0.032276 | 0.192238 | 0.005943 |
| 16585 | TMEM182 | 306 | 147434 | -0.13682 | 0.04977 | -0.23436 | -0.03927 | 0.005975 |
| 14239 | SEC14L3 | 609 | 147131 | 0.097051 | 0.03531 | 0.027839 | 0.166263 | 0.00599 |
| 12126 | PLAGL2 | 39 | 147701 | -0.38409 | 0.13994 | -0.65837 | -0.10981 | 0.006057 |
| 13886 | RPS6KB1 | 304 | 147436 | -0.13769 | 0.05019 | -0.23607 | -0.03932 | 0.006082 |
| 17042 | TRIM38 | 165 | 147575 | 0.186628 | 0.06807 | 0.053217 | 0.320039 | 0.00611 |
| 4276 | DIO1 | 278 | 147462 | -0.14295 | 0.05217 | -0.24521 | -0.0407 | 0.006143 |
| 147 | ACER2 | 308 | 147432 | 0.136556 | 0.04985 | 0.038856 | 0.234255 | 0.006154 |
| 5149 | EXOC8 | 374 | 147366 | -0.12392 | 0.04525 | -0.2126 | -0.03523 | 0.006169 |
| 1426 | BARHL1 | 442 | 147298 | 0.113962 | 0.04163 | 0.032366 | 0.195557 | 0.006192 |
| 5045 | ERI1 | 210 | 147530 | 0.164909 | 0.06034 | 0.04665 | 0.283168 | 0.006274 |
| 6725 | GYS2 | 1114 | 146626 | 0.071328 | 0.02611 | 0.02016 | 0.122496 | 0.006291 |
| 6204 | GIP | 137 | 147603 | 0.201274 | 0.07389 | 0.056458 | 0.346091 | 0.006448 |
| 10196 | NBPF6 | 393 | 147347 | 0.120211 | 0.04414 | 0.033701 | 0.206721 | 0.00646 |
| 6119 | GDF5 | 331 | 147409 | 0.122862 | 0.04512 | 0.034433 | 0.21129 | 0.006466 |
| 8963 | LTK | 2020 | 145720 | 0.052918 | 0.01947 | 0.014768 | 0.091069 | 0.006555 |
| 10064 | MYO5B | 2928 | 144812 | 0.043596 | 0.01607 | 0.0121 | 0.075092 | 0.006669 |
| 11094 | OR2T29 | 22 | 147718 | -0.50434 | 0.18628 | -0.86945 | -0.13923 | 0.006782 |
| 8858 | LRRC29 | 172 | 147568 | -0.18031 | 0.06667 | -0.31098 | -0.04963 | 0.006842 |
| 9402 | MEP1A | 717 | 147023 | 0.088013 | 0.03258 | 0.024158 | 0.151868 | 0.006903 |
| 14084 | SAR1A | 39 | 147701 | 0.378056 | 0.13994 | 0.103765 | 0.652346 | 0.006904 |
| 17989 | WNT8B | 213 | 147527 | 0.161854 | 0.05991 | 0.044425 | 0.279283 | 0.006904 |
| 53 | ABCC10 | 3192 | 144548 | 0.041362 | 0.01531 | 0.011349 | 0.071375 | 0.006911 |
| 18392 | ZNF232 | 110 | 147630 | 0.224927 | 0.08335 | 0.061569 | 0.388284 | 0.006961 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 11573 | PBDC1 | 51 | 147689 | 0.330139 | 0.12253 | 0.08999 | 0.570288 | 0.007051 |
| 17032 | TRIM29 | 471 | 147269 | 0.108113 | 0.04021 | 0.029309 | 0.186917 | 0.007168 |
| 17579 | UPK3A | 687 | 147053 | -0.08966 | 0.03335 | -0.15502 | -0.0243 | 0.00717 |
| 15040 | SLURP1 | 118 | 147622 | -0.21636 | 0.08047 | -0.37408 | -0.05863 | 0.007175 |
| 12977 | PTPN9 | 609 | 147131 | -0.09506 | 0.0354 | -0.16444 | -0.02568 | 0.007244 |
| 8784 | LPAR3 | 163 | 147577 | 0.183745 | 0.06849 | 0.049507 | 0.317983 | 0.007301 |
| 11820 | PER3 | 663 | 147077 | -0.09039 | 0.03371 | -0.15646 | -0.02432 | 0.007333 |
| 12823 | PSAT1 | 666 | 147074 | 0.090858 | 0.03395 | 0.024311 | 0.157406 | 0.007451 |
| 6256 | GLIS1 | 803 | 146937 | -0.08182 | 0.03058 | -0.14175 | -0.02188 | 0.007462 |
| 1828 | C17orf67 | 43 | 147697 | 0.355772 | 0.13328 | 0.094545 | 0.616999 | 0.0076 |
| 11038 | OR1L3 | 109 | 147631 | 0.223198 | 0.08375 | 0.059055 | 0.387341 | 0.007696 |
| 16632 | TMEM229 | 472 | 147268 | 0.10635 | 0.03991 | 0.028127 | 0.184572 | 0.007705 |
| 10223 | NCKAP5 | 472 | 147268 | 0.106279 | 0.03991 | 0.028056 | 0.184502 | 0.007746 |
| 6847 | HDX | 180 | 147560 | 0.172061 | 0.06466 | 0.045322 | 0.298801 | 0.007794 |
| 2246 | CARNS1 | 1811 | 145929 | 0.05463 | 0.02054 | 0.014368 | 0.094892 | 0.007828 |
| 16400 | TIPARP | 78 | 147662 | -0.26322 | 0.09898 | -0.45722 | -0.06922 | 0.007829 |
| 776 | ANXA11 | 531 | 147209 | -0.09918 | 0.03733 | -0.17234 | -0.02603 | 0.007877 |
| 5581 | FERD3L | 371 | 147369 | -0.11982 | 0.04513 | -0.20827 | -0.03136 | 0.007933 |
| 16747 | TMLHE | 552 | 147188 | -0.09845 | 0.03712 | -0.1712 | -0.02569 | 0.007998 |
| 7738 | ITGAV | 661 | 147079 | 0.089887 | 0.03391 | 0.023416 | 0.156357 | 0.008039 |
| 9541 | MIOX | 376 | 147364 | 0.118167 | 0.04461 | 0.030736 | 0.205599 | 0.008073 |
| 15396 | SPEM2 | 300 | 147440 | 0.133109 | 0.05026 | 0.034608 | 0.231611 | 0.008083 |
| 4425 | DNASE1 | 1152 | 146588 | -0.06752 | 0.02552 | -0.11754 | -0.0175 | 0.008155 |
| 7762 | ITM2B | 223 | 147517 | -0.15061 | 0.05704 | -0.26241 | -0.03881 | 0.008283 |
| 8441 | LAMA5 | 10011 | 137729 | -0.02266 | 0.00859 | -0.03949 | -0.00582 | 0.008343 |
| 15564 | SSMEM1 | 260 | 147480 | -0.14305 | 0.05424 | -0.24937 | -0.03673 | 0.008362 |
| 17751 | VIPR1 | 515 | 147225 | -0.09915 | 0.03761 | -0.17286 | -0.02545 | 0.008374 |
| 16875 | TOMM70 | 251 | 147489 | 0.142834 | 0.05424 | 0.036519 | 0.24915 | 0.008458 |
| 3320 | CNNM4 | 1073 | 146667 | 0.069605 | 0.02644 | 0.017778 | 0.121431 | 0.00848 |
| 7949 | KCNN4 | 597 | 147143 | -0.09288 | 0.03531 | -0.16209 | -0.02367 | 0.008534 |
| 14334 | SERINC4 | 586 | 147154 | -0.09486 | 0.03608 | -0.16557 | -0.02416 | 0.008549 |
| 9625 | MN1 | 2046 | 145694 | -0.04997 | 0.01901 | -0.08723 | -0.01271 | 0.008576 |
| 848 | APLP2 | 1263 | 146477 | 0.063163 | 0.02405 | 0.016033 | 0.110293 | 0.008621 |
| 8588 | LGR6 | 1168 | 146572 | 0.06681 | 0.02545 | 0.016926 | 0.116694 | 0.008665 |
| 14756 | SLC25A4 | 183 | 147557 | -0.16694 | 0.0636 | -0.2916 | -0.04228 | 0.008673 |
| 477 | AIG1 | 371 | 147369 | -0.11425 | 0.04355 | -0.19961 | -0.02889 | 0.008707 |
| 18555 | ZNF511 | 701 | 147039 | 0.072151 | 0.02751 | 0.018232 | 0.12607 | 0.008724 |
| 71 | ABCG1 | 893 | 146847 | -0.07556 | 0.02882 | -0.13205 | -0.01908 | 0.008745 |
| 16320 | THBS3 | 1910 | 145830 | 0.052521 | 0.02004 | 0.013249 | 0.091793 | 0.008762 |
| 15111 | SMIM5 | 76 | 147664 | -0.26277 | 0.10029 | -0.45933 | -0.0662 | 0.008791 |
| 15019 | SLFN12 | 615 | 147125 | -0.08926 | 0.03414 | -0.15616 | -0.02235 | 0.008931 |
| 6269 | GLRA4 | 460 | 147280 | -0.10632 | 0.04073 | -0.18615 | -0.02649 | 0.009047 |
| 5292 | FAM171A1 | 747 | 146993 | 0.083278 | 0.03193 | 0.020701 | 0.145855 | 0.009098 |
| 13902 | RRAGB | 128 | 147612 | -0.20159 | 0.07731 | -0.35311 | -0.05007 | 0.009115 |
| 11678 | PCNT | 2377 | 145363 | -0.04667 | 0.0179 | -0.08175 | -0.01159 | 0.009122 |
| 2824 | CEL | 4187 | 143553 | 0.031455 | 0.01207 | 0.007793 | 0.055118 | 0.009175 |
| 11616 | PCDHA8 | 538 | 147202 | 0.097288 | 0.03735 | 0.024076 | 0.1705 | 0.009201 |
| 9304 | MCM6 | 287 | 147453 | -0.13378 | 0.05137 | -0.23446 | -0.03309 | 0.009209 |
| 11073 | OR2H1 | 98 | 147642 | -0.22962 | 0.0883 | -0.40269 | -0.05655 | 0.009313 |
| 15611 | ST8SIA1 | 165 | 147575 | -0.17681 | 0.06807 | -0.31022 | -0.0434 | 0.00939 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 9499 | MIB1 | 969 | 146771 | -0.07211 | 0.02782 | -0.12663 | -0.01759 | 0.009538 |
| 5169 | EYA3 | 562 | 147178 | -0.09545 | 0.03684 | -0.16764 | -0.02325 | 0.009566 |
| 13518 | RGS4 | 254 | 147486 | -0.14223 | 0.0549 | -0.24983 | -0.03463 | 0.009573 |
| 16070 | TBL3 | 1960 | 145780 | 0.051125 | 0.01974 | 0.012444 | 0.089805 | 0.009583 |
| 14889 | SLC39A8 | 356 | 147384 | -0.11989 | 0.04637 | -0.21078 | -0.029 | 0.009726 |

**S2. Table 7.** Rare variant SAMHD1 phenome-wide association testing with disease status. Nominally significant (P<0.01) results are shown.

| Index | Phecode | Phenotype Description | Category | # Case | # Control | # Case w/ Mutation | # Case w/o Mutation | # Control w/ mutation | # Control w/o mutation | OR | 95% LCI | 95% UCI | P-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 44 | phecode_174 | Breast cancer | neoplasms | 6727 | 162518 | 86 | 6641 | 1154 | 161364 | 1.91 | 1.52 | 2.4 | 2.73E-08 |
| 46 | phecode_174.11 | Malignant neoplasm of female breast | neoplasms | 6186 | 84387 | 79 | 6107 | 566 | 83821 | 1.9 | 1.5 | 2.41 | 1.01E-07 |
| 45 | phecode_174.1 | Breast cancer [female] | neoplasms | 6631 | 84387 | 83 | 6548 | 566 | 83821 | 1.86 | 1.48 | 2.35 | 1.38E-07 |
| 65 | phecode_195.1 | Malignant neoplasm- other | neoplasms | 14784 | 151547 | 159 | 14625 | 1053 | 150494 | 1.54 | 1.3 | 1.82 | 5.78E-07 |
| 64 | phecode_195 | Cancer- suspected or other | neoplasms | 15170 | 151547 | 161 | 15009 | 1053 | 150494 | 1.52 | 1.28 | 1.8 | 1.05E-06 |
| 66 | phecode_197 | Chemotherapy | neoplasms | 9299 | 151547 | 105 | 9194 | 1053 | 150494 | 1.61 | 1.32 | 1.98 | 3.16E-06 |
| 21 | phecode_1010 | Other tests | NULL | 9440 | 164248 | 103 | 9337 | 1173 | 163075 | 1.52 | 1.24 | 1.86 | 5.39E-05 |
| 47 | phecode_175 | Acquired absence of breast | neoplasms | 1269 | 162044 | 21 | 1248 | 1148 | 160896 | 2.42 | 1.57 | 3.74 | 7.02E-05 |
| 630 | phecode_702.2 | Seborrheic keratosis | dermatologic | 1422 | 171025 | 24 | 1398 | 1247 | 169778 | 2.18 | 1.46 | 3.27 | 0.000148 |
| 86 | phecode_214 | Lipoma | neoplasms | 2996 | 170201 | 38 | 2958 | 1229 | 168972 | 1.78 | 1.29 | 2.45 | 0.000384 |
| 56 | phecode_189 | Cancer of urinary organs (incl. kidney and bladder) | neoplasms | 2182 | 171506 | 31 | 2151 | 1245 | 170261 | 1.91 | 1.34 | 2.74 | 0.000399 |
| 67 | phecode_198 | Secondary malignant neoplasm | neoplasms | 5038 | 151547 | 55 | 4983 | 1053 | 150494 | 1.58 | 1.21 | 2.08 | 0.000933 |
| 524 | phecode_599.3 | Dysuria | genitourinary | 560 | 161318 | 11 | 549 | 1160 | 160158 | 2.65 | 1.47 | 4.79 | 0.001252 |
| 7 | phecode_041 | Bacterial infection NOS | infectious diseases | 5860 | 166072 | 65 | 5795 | 1201 | 164871 | 1.49 | 1.16 | 1.91 | 0.001792 |
| 73 | phecode_198.6 | Secondary malignancy of bone | neoplasms | 1236 | 151547 | 18 | 1218 | 1053 | 150494 | 2.1 | 1.31 | 3.35 | 0.001875 |
| 648 | phecode_717 | Polymyalgia Rheumatica | musculoskeletal | 665 | 173023 | 12 | 653 | 1264 | 171759 | 2.47 | 1.39 | 4.37 | 0.002002 |
| 633 | phecode_704 | Diseases of hair and hair follicles | dermatologic | 2459 | 170744 | 31 | 2428 | 1238 | 169506 | 1.72 | 1.21 | 2.46 | 0.00269 |
| 9 | phecode_041.2 | Streptococcus infection | infectious diseases | 799 | 166072 | 13 | 786 | 1201 | 164871 | 2.26 | 1.31 | 3.9 | 0.003359 |
| 721 | phecode_793 | Nonspecific abnormal findings on radiological and other examination of musculoskeletal system | symptoms | 377 | 173311 | 7 | 370 | 1269 | 172042 | 2.79 | 1.4 | 5.56 | 0.003545 |
| 544 | phecode_611.3 | Lump or mass in breast | genitourinary | 722 | 92533 | 12 | 710 | 653 | 91880 | 2.33 | 1.32 | 4.13 | 0.003712 |
| 544 | phecode_611.3 | Lump or mass in breast | genitourinary | 722 | 92533 | 12 | 710 | 653 | 91880 | 2.33 | 1.32 | 4.13 | 0.003712 |
| 80 | phecode_204.1 | Lymphoid leukemia | neoplasms | 397 | 171046 | 8 | 389 | 1247 | 169799 | 2.79 | 1.39 | 5.58 | 0.003752 |
| 79 | phecode_204 | Leukemia | neoplasms | 1068 | 171046 | 16 | 1052 | 1247 | 169799 | 2.07 | 1.27 | 3.39 | 0.003792 |
| 58 | phecode_189.11 | Malignant neoplasm of kidney- except pelvis | neoplasms | 643 | 171506 | 11 | 632 | 1245 | 170261 | 2.32 | 1.28 | 4.2 | 0.005398 |
| 87 | phecode_214.1 | Lipoma of skin and subcutaneous tissue | neoplasms | 2269 | 170201 | 27 | 2242 | 1229 | 168972 | 1.7 | 1.17 | 2.47 | 0.0054 |
| 543 | phecode_611 | Abnormal findings on mammogram or breast exam | genitourinary | 748 | 92533 | 12 | 736 | 653 | 91880 | 2.25 | 1.27 | 3.99 | 0.005503 |
| 472 | phecode_574 | Cholelithiasis and cholecystitis | digestive | 7413 | 165528 | 75 | 7338 | 1197 | 164331 | 1.38 | 1.1 | 1.75 | 0.006495 |
| 57 | phecode_189.1 | Cancer of kidney and renal pelvis | neoplasms | 666 | 171506 | 11 | 655 | 1245 | 170261 | 2.24 | 1.24 | 4.05 | 0.007689 |
| 88 | phecode_215 | Other benign neoplasm of connective and other soft tissue | neoplasms | 508 | 170201 | 9 | 499 | 1229 | 168972 | 2.43 | 1.26 | 4.67 | 0.00803 |
| 36 | phecode_159 | Malignant neoplasm of other and ill-defined sites within the digestive organs and peritoneum | neoplasms | 2889 | 165054 | 34 | 2855 | 1199 | 163855 | 1.57 | 1.12 | 2.22 | 0.009233 |
| 610 | phecode_686 | Other local infections of skin and subcutaneous tissue | dermatologic | 2115 | 168268 | 26 | 2089 | 1224 | 167044 | 1.67 | 1.13 | 2.46 | 0.00943 |
| 335 | phecode_446 | Polyarteritis nodosa and allied conditions | circulatory system | 435 | 169127 | 8 | 427 | 1241 | 167886 | 2.49 | 1.25 | 4.98 | 0.009768 |
| 521 | phecode_599 | Other symptoms/disorders or the urinary system | genitourinary | 12370 | 161318 | 116 | 12254 | 1160 | 160158 | 1.29 | 1.06 | 1.56 | 0.009848 |

**S2. Table 8**. Phenoscanner search results for genetic variants initially considered as genetic instruments for Mendelian Randomization analysis.

| Index | rsid | Variant Excluded from MR? | Reason for Inclusion or Exclusion | Gene | Trait | study | pmid | ancestry | year | p | n |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | rs12148 | No | Coding mutation in Mitochondrially Localized Gene | *SCO2* | Mean corpuscular hemoglobin | Astle W | 27863252 | European | 2016 | 7.54E-85 | 173480 |
| 2 | rs12148 | No | Coding mutation in Mitochondrially Localized Gene | *SCO2* | Mean corpuscular volume | Astle W | 27863252 | European | 2016 | 3.63E-113 | 173480 |
| 3 | rs12148 | No | Coding mutation in Mitochondrially Localized Gene | *SCO2* | Red blood cell count | Astle W | 27863252 | European | 2016 | 3.84E-38 | 173480 |
| 4 | rs12148 | No | Coding mutation in Mitochondrially Localized Gene | *SCO2* | Red cell distribution width | Astle W | 27863252 | European | 2016 | 7.83E-19 | 173480 |
| 5 | rs12148 | No | Coding mutation in Mitochondrially Localized Gene | *SCO2* | Reticulocyte count | Astle W | 27863252 | European | 2016 | 2.08E-14 | 173480 |
| 6 | rs12148 | No | Coding mutation in Mitochondrially Localized Gene | *SCO2* | Mean corpuscular hemoglobin concentration | van der Harst P | 23222517 | Mixed | 2012 | 5.64E-13 | 71861 |
| 7 | rs12148 | No | Coding mutation in Mitochondrially Localized Gene | *SCO2* | Mean corpuscular volume | Ganesh SK | 19862010 | European | 2009 | 1.23E-09 | 24167 |
| 8 | rs12148 | No | Coding mutation in Mitochondrially Localized Gene | *SCO2* | Mean corpuscular volume | van der Harst P | 23222517 | Mixed | 2012 | 4.41E-14 | 71861 |
| 9 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Granulocyte percentage of myeloid white cells | Astle W | 27863252 | European | 2016 | 2.62E-09 | 173480 |
| 10 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Neutrophil percentage of white cells | Astle W | 27863252 | European | 2016 | 8.90E-09 | 173480 |
| 11 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Arm fat-free mass left | Neale B | UKBB | European | 2017 | 1.10E-19 | 331159 |
| 12 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Arm fat-free mass right | Neale B | UKBB | European | 2017 | 1.27E-20 | 331221 |
| 13 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Arm predicted mass left | Neale B | UKBB | European | 2017 | 4.32E-20 | 331146 |
| 14 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Arm predicted mass right | Neale B | UKBB | European | 2017 | 2.42E-21 | 331216 |
| 15 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Basal metabolic rate | Neale B | UKBB | European | 2017 | 1.16E-20 | 331307 |
| 16 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Comparative height size at age 10 | Neale B | UKBB | European | 2017 | 1.17E-32 | 332021 |
| 17 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Forced expiratory volume in 1-second, predicted | Neale B | UKBB | European | 2017 | 4.86E-15 | 110423 |
| 18 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Height | Neale B | UKBB | European | 2017 | 6.14E-55 | 336474 |
| 19 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Leg fat-free mass left | Neale B | UKBB | European | 2017 | 1.29E-14 | 331258 |
| 20 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Leg fat-free mass right | Neale B | UKBB | European | 2017 | 4.48E-16 | 331285 |
| 21 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Leg predicted mass left | Neale B | UKBB | European | 2017 | 1.08E-14 | 331253 |
| 22 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Leg predicted mass right | Neale B | UKBB | European | 2017 | 2.97E-16 | 331285 |
| 23 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Trunk fat-free mass | Neale B | UKBB | European | 2017 | 3.63E-26 | 331030 |
| 24 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Trunk predicted mass | Neale B | UKBB | European | 2017 | 1.35E-26 | 330995 |
| 25 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Weight | Neale B | UKBB | European | 2017 | 6.16E-13 | 336227 |
| 26 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Whole body fat-free mass | Neale B | UKBB | European | 2017 | 7.00E-23 | 331291 |
| 27 | rs8067252 | Yes | Strong associations with anthropometric measurements | *ADAP2* | Whole body water mass | Neale B | UKBB | European | 2017 | 2.19E-22 | 331315 |
| 28 | rs56069439 | Yes | Strong association with female cancers | *ANKLE1* | Breast cancer risk in BRCA1 mutation carriers | Antoniou AC | 20852631 | European | 2010 | 1.06E-11 | 2383 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 29 | rs56069439 | Yes | Strong association with female cancers | *ANKLE1* | Breast cancer | Couch FJ | 27117709 | European | 2016 | 1.00E-32 | - |
| 30 | rs56069439 | Yes | Strong association with female cancers | *ANKLE1* | Breast cancer | Michailidou K | 29059683 | Mixed | 2017 | 9.00E-09 | - |
| 31 | rs56069439 | Yes | Strong association with female cancers | *ANKLE1* | Breast cancer estrogen receptor negative | Couch FJ | 27117709 | European | 2016 | 8.00E-19 | - |
| 32 | rs56069439 | Yes | Strong association with female cancers | *ANKLE1* | High grade serous ovarian cancer | Phelan M | 28346442 | European | 2017 | 6.48E-26 | 53978 |
| 33 | rs56069439 | Yes | Strong association with female cancers | *ANKLE1* | Invasive ovarian cancer | Phelan M | 28346442 | European | 2017 | 2.94E-17 | 63347 |
| 34 | rs56069439 | Yes | Strong association with female cancers | *ANKLE1* | Serous invasive ovarian cancer | Phelan M | 28346442 | European | 2017 | 6.66E-24 | 54990 |
| 35 | rs701834 | No | Minimal evidence of pleiotropy | *LZTS2* | Comparative height size at age 10 | Neale B | UKBB | European | 2017 | 1.39E-12 | 332021 |
| 36 | rs12426673 | No | Coding mutation in Mitochondrially Localized Gene | *USP30-AS1* | Platelet count | Astle W | 27863252 | European | 2016 | 4.15E-31 | 173480 |
| 37 | rs12426673 | No | Coding mutation in Mitochondrially Localized Gene | *USP30-AS1* | Plateletcrit | Astle W | 27863252 | European | 2016 | 4.57E-39 | 173480 |
| 38 | rs17850455 | No | Coding mutation in Mitochondrially Localized Gene | *POLG2* | Cause of death: unspecified place | Neale B | UKBB | European | 2017 | 2.66E-08 | 7637 |
| 39 | rs5745582 | No | Coding mutation in Mitochondrially Localized Gene | *BAK1* | Platelet count | Astle W | 27863252 | European | 2016 | 3.36E-101 | 173480 |
| 40 | rs5745582 | No | Coding mutation in Mitochondrially Localized Gene | *BAK1* | Plateletcrit | Astle W | 27863252 | European | 2016 | 2.72E-116 | 173480 |
| 41 | rs5745582 | No | Coding mutation in Mitochondrially Localized Gene | *BAK1* | Platelet count | Qayyum R | 22423221 | African | 2012 | 2.48E-12 | 16388 |
| 42 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Mean platelet volume | Astle W | 27863252 | European | 2016 | 1.79E-09 | 173480 |
| 43 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | HIV 1 control | Pereyra F | 21051598 | Mixed | 2010 | 1.20E-10 | 3622 |
| 44 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | HIV 1 control viral load at set point | Fellay J | 20041166 | European | 2009 | 6.40E-09 | 2362 |
| 45 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Rheumatoid arthritis | Plenge RM | 17804836 | European | 2007 | 8.49E-16 | 3324 |
| 46 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Rheumatoid arthritis | Gregersen PK | 19503088 | European | 2009 | 5.52E-27 | 6922 |
| 47 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Rheumatoid arthritis | Stahl E | 20453842 | European | 2010 | 1.47E-46 | 25708 |
| 48 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Ulcerative colitis | IBDGC | 26192919 | European | 2015 | 1.21E-10 | 27432 |
| 49 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Primary sclerosing cholangitis | Ji S | 27992413 | European | 2017 | 1.36E-27 | 14890 |
| 50 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Psoriasis | Neale B | UKBB | European | 2017 | 3.60E-09 | 337199 |
| 51 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Self-reported hyperthyroidism or thyrotoxicosis | Neale B | UKBB | European | 2017 | 1.48E-14 | 337159 |
| 52 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Self-reported psoriasis | Neale B | UKBB | European | 2017 | 3.08E-115 | 337159 |
| 53 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Self-reported rheumatoid arthritis | Neale B | UKBB | European | 2017 | 6.22E-10 | 337159 |
| 54 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Treatment with dovobet ointment | Neale B | UKBB | European | 2017 | 8.51E-15 | 337159 |
| 55 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Treatment with dovonex 50micrograms or g cream | Neale B | UKBB | European | 2017 | 8.39E-09 | 337159 |
| 56 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Treatment with levothyroxine sodium | Neale B | UKBB | European | 2017 | 1.13E-10 | 337159 |
| 57 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Rheumatoid arthritis | Okada Y | 24390342 | East Asian | 2014 | 1.40E-14 | 22515 |
| 57 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Rheumatoid arthritis | Okada Y | 24390342 | East Asian | 2014 | 1.40E-14 | 22515 |
| 58 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Rheumatoid arthritis | Okada Y | 24390342 | European | 2014 | 5.10E-82 | 58284 |
| 59 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Rheumatoid arthritis | Okada Y | 24390342 | Mixed | 2014 | 6.00E-32 | 80799 |
| 60 | rs2844509 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Rheumatoid arthritis | Stahl E | 20453842 | European | 2010 | 1.47E-46 | 25706 |
| 61 | rs73004962 | Yes | Strong associations with immune diseases | *ATP6V1G2-DDX39B* | Lymphocyte percentage of white cells | Astle W | 27863252 | European | 2016 | 1.17E-08 | 173480 |
| 62 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Plateletcrit | Astle W | 27863252 | European | 2016 | 2.21E-13 | 173480 |
| 63 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Type II diabetes | DIAGRAM | 26551672 | European | 2015 | 1.10E-09 | 84780 |
| 64 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Low density lipoprotein | GLGC | 24097068 | European | 2013 | 2.71E-21 | 74969 |
| 65 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Total cholesterol | GLGC | 24097068 | European | 2013 | 7.94E-23 | 82981 |
| 66 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Triglycerides | GLGC | 24097068 | European | 2013 | 6.65E-40 | 78333 |
| 67 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Height | Neale B | UKBB | European | 2017 | 7.37E-11 | 336474 |
| 68 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Leg fat-free mass left | Neale B | UKBB | European | 2017 | 2.82E-08 | 331258 |
| 69 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Leg fat-free mass right | Neale B | UKBB | European | 2017 | 1.52E-08 | 331285 |
| 70 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Leg predicted mass left | Neale B | UKBB | European | 2017 | 2.41E-08 | 331253 |
| 71 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Leg predicted mass right | Neale B | UKBB | European | 2017 | 1.97E-08 | 331285 |
| 72 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Medication for cholesterol, blood pressure or diabetes: cholesterol lowering medication | Neale B | UKBB | European | 2017 | 1.06E-12 | 154702 |
| 73 | rs73004962 | Yes | Strong associations with lipids | *PBX4* | Self-reported high cholesterol | Neale B | UKBB | European | 2017 | 1.64E-14 | 337159 |
| 74 | rs1760940 | No | Minimal evidence of pleiotropy and has a known role in mitochondrial function (https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007654) | *PNP* | Age at menopause | ReproGen | 26414677 | European | 2015 | 5.10E-10 | 69360 |
| 75 | rs3766744 | No | Coding mutation in Mitochondrially Localized Gene | *MFN2* | Platelet count | Astle W | 27863252 | European | 2016 | 5.02E-25 | 173480 |
| 76 | rs3766744 | No | Coding mutation in Mitochondrially Localized Gene | *MFN2* | Plateletcrit | Astle W | 27863252 | European | 2016 | 2.11E-37 | 173480 |
| 77 | rs6792510 | No | Minimal evidence of pleiotropy | *NCKIPSD* | Sum eosinophil basophil counts | Astle W | 27863252 | European | 2016 | 1.93E-08 | 173480 |
| 78 | rs11596235 | No | Minimal evidence of pleiotropy | *SUFU* | Sum eosinophil basophil counts | Astle W | 27863252 | European | 2016 | 3.91E-08 | 173480 |
| 79 | rs11596235 | No | Minimal evidence of pleiotropy | *SUFU* | Waist circumference adjusted for BMI | GIANT | 25673412 | Mixed | 2015 | 1.50E-08 | 243954 |
| 80 | rs11596235 | No | Minimal evidence of pleiotropy | *SUFU* | Height | Neale B | UKBB | European | 2017 | 1.40E-11 | 336474 |
| 81 | rs11596235 | No | Minimal evidence of pleiotropy | *SUFU* | Sitting height | Neale B | UKBB | European | 2017 | 4.83E-10 | 336172 |
| 82 | rs1342442 | No | Minimal evidence of pleiotropy and experimental evidence suggesting a direct role in mtDNA regulation (https://pubmed.ncbi.nlm.nih.gov/21393861/) | *MEF2D* | Platelet count | Astle W | 27863252 | European | 2016 | 7.97E-11 | 173480 |
| 83 | rs1342442 | No | Minimal evidence of pleiotropy and experimental evidence suggesting a direct role in mtDNA regulation (https://pubmed.ncbi.nlm.nih.gov/21393861/) | *MEF2D* | Plateletcrit | Astle W | 27863252 | European | 2016 | 3.06E-14 | 173480 |
| 84 | rs1342442 | No | Minimal evidence of pleiotropy and experimental evidence suggesting a direct role in mtDNA regulation (https://pubmed.ncbi.nlm.nih.gov/21393861/) | *MEF2D* | Medication for pain relief, constipation, heartburn: none of the above | Neale B | UKBB | European | 2017 | 3.45E-09 | 333581 |
| 85 | rs1342442 | No | Minimal evidence of pleiotropy and experimental evidence suggesting a direct role in mtDNA regulation (https://pubmed.ncbi.nlm.nih.gov/21393861/) | *MEF2D* | Medication for pain relief, constipation, heartburn: paracetamol | Neale B | UKBB | European | 2017 | 8.13E-09 | 333581 |
| 86 | rs1342442 | No | Minimal evidence of pleiotropy and experimental evidence suggesting a direct role in mtDNA regulation (https://pubmed.ncbi.nlm.nih.gov/21393861/) | *MEF2D* | Pain type experienced in last month: headache | Neale B | UKBB | European | 2017 | 8.12E-16 | 336650 |
| 87 | rs1342442 | No | Minimal evidence of pleiotropy and experimental evidence suggesting a direct role in mtDNA regulation (https://pubmed.ncbi.nlm.nih.gov/21393861/) | *MEF2D* | Pulse rate | Neale B | UKBB | European | 2017 | 3.01E-12 | 317756 |
| 88 | rs1342442 | No | Minimal evidence of pleiotropy and experimental evidence suggesting a direct role in mtDNA regulation (https://pubmed.ncbi.nlm.nih.gov/21393861/) | *MEF2D* | Self-reported migraine | Neale B | UKBB | European | 2017 | 1.10E-08 | 337159 |
| 89 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | High light scatter percentage of red cells | Astle W | 27863252 | European | 2016 | 2.95E-38 | 173480 |
| 90 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | High light scatter reticulocyte count | Astle W | 27863252 | European | 2016 | 9.58E-36 | 173480 |
| 91 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Immature fraction of reticulocytes | Astle W | 27863252 | European | 2016 | 4.36E-38 | 173480 |
| 92 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Red cell distribution width | Astle W | 27863252 | European | 2016 | 1.97E-44 | 173480 |
| 93 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Reticulocyte count | Astle W | 27863252 | European | 2016 | 8.88E-18 | 173480 |
| 94 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Reticulocyte fraction of red cells | Astle W | 27863252 | European | 2016 | 4.40E-20 | 173480 |
| 95 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Coronary artery disease | CARDIoGRAMplusC4D | 26343387 | Mixed | 2015 | 8.17E-11 | 184305 |
| 96 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | High density lipoprotein | GLGC | 24097068 | European | 2013 | 4.44E-19 | 92158 |
| 97 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Low density lipoprotein | GLGC | 24097068 | European | 2013 | 1.24e-652 | 82533 |
| 98 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Total cholesterol | GLGC | 24097068 | European | 2013 | 1.56E-283 | 92046 |
| 99 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Triglycerides | GLGC | 24097068 | European | 2013 | 1.15E-28 | 86164 |
| 100 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | APOB apolipoprotein B | Hopewell | 23100282 | European | 2012 | 5.90E-154 | 3895 |
| 101 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | APOB apolipoprotein B response after 40mg daily simvastatin treatment | Hopewell | 23100282 | European | 2012 | 4.90E-29 | 3895 |
| 102 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol | Smith EN | 20838585 | European | 2010 | 1.60E-08 | 525 |
| 103 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol | Asselbergs | 23063622 | European | 2012 | 1.52E-71 | 66240 |
| 104 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol | Rasmussen Torvik LJ | 23067351 | African | 2012 | 1.50E-09 | 1249 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 104 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol | Rasmussen Torvik LJ | 23067351 | African | 2012 | 1.50E-09 | 1249 |
| 105 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol | Hopewell | 23100282 | European | 2012 | 2.10E-215 | 3895 |
| 106 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol among a group where cancer diabetes hyperhypothyroidism and LDL altering medications were not present | Rasmussen Torvik LJ | 23067351 | African | 2012 | 6.30E-11 | 1249 |
| 107 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol change with statins | Thompson | 20031582 | Mixed | 2009 | 5.54E-30 | 1984 |
| 108 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol response after 40mg daily simvastatin treatment | Hopewell | 23100282 | European | 2012 | 2.70E-30 | 3895 |
| 109 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol response to statins baseline LDL cholesterol | Chasman DI | 22331829 | European | 2012 | 1.60E-47 | 6989 |
| 110 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol response to statins fractional change in LDL cholesterol | Chasman DI | 22331829 | European | 2012 | 5.80E-19 | 6989 |
| 111 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol response to statins residuals of the measure of fractional change in LDL cholesterol | Chasman DI | 22331829 | European | 2012 | 4.00E-17 | 6989 |
| 112 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Late onset Alzheimers disease | Naj AC | 20885792 | Unspecifie | 2010 | 5.49E-58 | 2035 |
| 113 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Late onset Alzheimers disease | Hu X | 21390209 | European | 2011 | 3.93E-09 | 3595 |
| 114 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Lipoprotein associated phospholipase A2 activity Lp2 | Chu AY | 23118302 | European | 2012 | 4.30E-81 | 6851 |
| 115 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Total cholesterol | Asselbergs | 23063622 | European | 2012 | 1.55E-36 | 66240 |
| 116 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Alzheimers disease | IGAP | 24162737 | European | 2013 | 1.23E-22 | 54162 |
| 117 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Cholesterol total | Surakka I | 25961943 | European | 2015 | 8.00E-239 | . |
| 118 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Cholesterol total | Nagy R | 28270201 | European | 2017 | 5.00E-94 | . |
| 119 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Coronary artery disease | Nelson CP | 28714975 | Mixed | 2017 | 2.00E-19 | . |
| 120 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Coronary artery disease | van der Harst P | 29212778 | Mixed | 2018 | 2.00E-35 | . |
| 121 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Coronary artery disease | van der Harst P | 29212778 | Mixed | 2018 | 3.00E-39 | . |
| 122 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Coronary artery disease | van der Harst P | 29212778 | Mixed | 2018 | 3.00E-39 | . |
| 123 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | HDL cholesterol | Nagy R | 28270201 | European | 2017 | 6.00E-14 | . |
| 124 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | High light scatter reticulocyte count | Astle W | 27863252 | European | 2016 | 1.00E-35 | . |
| 125 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | High light scatter reticulocyte percentage of red cells | Astle W | 27863252 | European | 2016 | 3.00E-38 | . |
| 126 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Ideal cardiovascular health clinical and behavioural | Allen NB | 27179730 | European | 2016 | 9.00E-16 | . |
| 127 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Immature fraction of reticulocytes | Astle W | 27863252 | European | 2016 | 4.00E-38 | . |
| 128 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol | Rasmussen Torvik LJ | 23067351 | African | 2012 | 2.00E-09 | . |
| 129 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol levels | Spracklen CN | 28334899 | East Asian | 2017 | 2.00E-286 | . |
| 130 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | LDL cholesterol levels | Zhu Y | 28371326 | East Asian | 2017 | 7.00E-15 | . |
| 131 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Lipid metabolism phenotypes | Kettunen J | 22286219 | European | 2012 | 3.00E-58 | . |
| 132 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Lipid traits | Wu Y | 24023260 | Filipino | 2013 | 2.00E-30 | . |
| 133 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Lipid traits | Wu Y | 24023260 | Filipino | 2013 | 3.00E-53 | . |
| 134 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Lipoprotein a levels | Mack S | 28512139 | European | 2017 | 3.00E-10 | . |
| 135 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Lipoprotein associated phospholipase A2 activity change in response to darapladib treatment in cardiovascular disease | Yeo A | 28753643 | Mixed | 2017 | 2.00E-27 | . |
| 136 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Lipoprotein phospholipase A2 activity in cardiovascular disease | Yeo A | 28753643 | Mixed | 2017 | 8.00E-78 | . |
| 137 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Lipoproteina levels adjusted for apolipoproteina isoforms | Mack S | 28512139 | European | 2017 | 3.00E-09 | . |
| 138 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Low density lipoprotein cholesterol | Southam L | 28548082 | Mixed | 2017 | 3.00E-19 | . |
| 139 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Metabolite levels lipoprotein measures | Kettunen J | 27005778 | European | 2016 | 2.00E-120 | . |
| 140 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Pulse pressure | Warren HR | 28135244 | European | 2017 | 4.00E-10 | . |
| 141 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Red cell distribution width | Astle W | 27863252 | European | 2016 | 2.00E-44 | . |
| 142 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Response to statin therapy LDL C | Chasman DI | 22331829 | European | 2012 | 2.00E-47 | . |
| 143 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Reticulocyte count | Astle W | 27863252 | European | 2016 | 9.00E-18 | . |
| 144 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Reticulocyte fraction of red cells | Astle W | 27863252 | European | 2016 | 4.00E-20 | . |
| 145 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Total cholesterol levels | Spracklen CN | 28334899 | East Asian | 2017 | 99999773478 | . |
| 146 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Total cholesterol levels | Southam L | 28548082 | Mixed | 2017 | 1.00E-08 | . |
| 147 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Arm fat mass left | Neale B | UKBB | European | 2017 | 2.01E-08 | 331164 |
| 148 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Arm fat mass right | Neale B | UKBB | European | 2017 | 5.01E-09 | 331226 |
| 149 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Arm fat percentage left | Neale B | UKBB | European | 2017 | 1.83E-08 | 331198 |
| 150 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Arm fat percentage right | Neale B | UKBB | European | 2017 | 1.61E-08 | 331249 |
| 151 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Body fat percentage | Neale B | UKBB | European | 2017 | 1.13E-09 | 331117 |
| 152 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Chronic ischaemic heart disease | Neale B | UKBB | European | 2017 | 1.43E-08 | 337199 |
| 153 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Hip circumference | Neale B | UKBB | European | 2017 | 2.68E-09 | 336601 |
| 154 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Illnesses of father: alzheimers disease or dementia | Neale B | UKBB | European | 2017 | 7.29E-15 | 292807 |
| 155 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Illnesses of father: heart disease | Neale B | UKBB | European | 2017 | 9.03E-09 | 298237 |
| 156 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Illnesses of mother: alzheimers disease or dementia | Neale B | UKBB | European | 2017 | 1.29E-40 | 309843 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 157 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Medication for cholesterol, blood pressure or diabetes: cholesterol lowering medication | Neale B | UKBB | European | 2017 | 2.85E-62 | 154702 |
| 158 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Medication for cholesterol, blood pressure or diabetes: none of the above | Neale B | UKBB | European | 2017 | 4.34E-18 | 154702 |
| 159 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Medication for pain relief, constipation, heartburn: aspirin | Neale B | UKBB | European | 2017 | 2.30E-08 | 333581 |
| 160 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | No treatment with medication for cholesterol, blood pressure, diabetes, or take exogenous hormones | Neale B | UKBB | European | 2017 | 1.12E-21 | 180203 |
| 161 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Number of treatments or medications taken | Neale B | UKBB | European | 2017 | 2.38E-08 | 337159 |
| 162 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Pulse rate | Neale B | UKBB | European | 2017 | 5.50E-17 | 317756 |
| 163 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Self-reported angina | Neale B | UKBB | European | 2017 | 1.38E-12 | 337159 |
| 164 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Self-reported heart attack or myocardial infarction | Neale B | UKBB | European | 2017 | 3.48E-09 | 337159 |
| 165 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Self-reported high cholesterol | Neale B | UKBB | European | 2017 | 4.73E-145 | 337159 |
| 166 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Treatment with aspirin | Neale B | UKBB | European | 2017 | 5.45E-10 | 337159 |
| 167 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Treatment with atorvastatin | Neale B | UKBB | European | 2017 | 4.91E-36 | 337159 |
| 168 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Treatment with cholesterol lowering medication | Neale B | UKBB | European | 2017 | 2.09E-86 | 180203 |
| 169 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Treatment with ezetimibe | Neale B | UKBB | European | 2017 | 1.05E-11 | 337159 |
| 170 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Treatment with lipitor 10mg tablet | Neale B | UKBB | European | 2017 | 1.69E-11 | 337159 |
| 171 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Treatment with rosuvastatin | Neale B | UKBB | European | 2017 | 1.72E-10 | 337159 |
| 172 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Treatment with simvastatin | Neale B | UKBB | European | 2017 | 5.02E-70 | 337159 |
| 173 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Trunk fat mass | Neale B | UKBB | European | 2017 | 7.96E-11 | 331093 |
| 174 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Trunk fat percentage | Neale B | UKBB | European | 2017 | 1.02E-10 | 331113 |
| 175 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Vascular or heart problems diagnosed by doctor: angina | Neale B | UKBB | European | 2017 | 1.28E-14 | 336683 |
| 176 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Vascular or heart problems diagnosed by doctor: heart attack | Neale B | UKBB | European | 2017 | 1.29E-08 | 336683 |
| 177 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Vascular or heart problems diagnosed by doctor: none of the above | Neale B | UKBB | European | 2017 | 1.40E-09 | 336683 |
| 178 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Weight | Neale B | UKBB | European | 2017 | 1.25E-08 | 336227 |
| 179 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Whole body fat mass | Neale B | UKBB | European | 2017 | 6.13E-10 | 330762 |
| 180 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Coronary artery disease | Nelson CP | 28714975 | Mixed | 2017 | 2.17E-19 | 148715 |
| 181 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Low density lipoprotein | Prins B | 28887542 | European | 2017 | 1.15E-77 | 9961 |
| 182 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Total cholesterol | Prins B | 28887542 | European | 2017 | 8.50E-32 | 9961 |
| 183 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | High density lipoprotein | Spracklen CN | 28334899 | East Asian | 2017 | 1.44E-08 | 10133 |
| 184 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Low density lipoprotein | Spracklen CN | 28334899 | East Asian | 2017 | 8.20E-107 | 10133 |
| 185 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Total cholesterol | Spracklen CN | 28334899 | East Asian | 2017 | 1.23E-40 | 10133 |
| 186 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Cholesterol ldl | Chasman DI | 22331829 | European | 2012 | 2.00E-47 | . |
| 187 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Lipid metabolism | Kettunen J | 22286219 | European | 2012 | 3.00E-58 | . |
| 188 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Coronary artery disease | van der Harst P | 29212778 | Mixed | 2018 | 7.30E-27 | 296525 |
| 189 | rs7412 | Yes | Strong associations with lipids and neurodegenerative disease | *APOE* | Coronary artery disease | van der Harst P | 29212778 | Mixed | 2018 | 2.14E-35 | 547261 |
| 190 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Granulocyte count | Astle W | 27863252 | European | 2016 | 1.48E-14 | 173480 |
| 191 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Granulocyte percentage of myeloid white cells | Astle W | 27863252 | European | 2016 | 3.07E-10 | 173480 |
| 192 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Lymphocyte percentage of white cells | Astle W | 27863252 | European | 2016 | 4.13E-12 | 173480 |
| 193 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Mean corpuscular hemoglobin | Astle W | 27863252 | European | 2016 | 8.84E-21 | 173480 |
| 194 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Mean corpuscular volume | Astle W | 27863252 | European | 2016 | 2.73E-18 | 173480 |
| 195 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Myeloid white cell count | Astle W | 27863252 | European | 2016 | 1.74E-13 | 173480 |
| 196 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Neutrophil count | Astle W | 27863252 | European | 2016 | 3.81E-13 | 173480 |
| 197 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Neutrophil percentage of white cells | Astle W | 27863252 | European | 2016 | 8.39E-12 | 173480 |
| 198 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Platelet count | Astle W | 27863252 | European | 2016 | 5.55E-184 | 173480 |
| 199 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Plateletcrit | Astle W | 27863252 | European | 2016 | 8.42E-238 | 173480 |
| 200 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Red blood cell count | Astle W | 27863252 | European | 2016 | 5.75E-15 | 173480 |
| 201 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Red cell distribution width | Astle W | 27863252 | European | 2016 | 5.71E-10 | 173480 |
| 202 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Sum basophil neutrophil counts | Astle W | 27863252 | European | 2016 | 1.21E-13 | 173480 |
| 203 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Sum neutrophil eosinophil counts | Astle W | 27863252 | European | 2016 | 4.26E-14 | 173480 |
| 204 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | White blood cell count | Astle W | 27863252 | European | 2016 | 9.43E-09 | 173480 |
| 205 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Platelet count PLT | Kamatani Y | 20139978 | East Asian | 2010 | 2.95E-13 | 14700 |
| 206 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Platelet count PLT 109l | Soranzo N | 19820697 | European | 2009 | 8.50E-17 | 4627 |
| 207 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Granulocyte count | Astle W | 27863252 | European | 2016 | 1.00E-14 | . |
| 208 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Granulocyte percentage of myeloid white cells | Astle W | 27863252 | European | 2016 | 3.00E-10 | . |

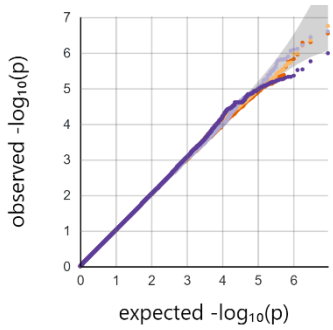
| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 209 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Hematological parameters | Soranzo N | 19820697 | European | 2009 | 9.00E-17 | . |
| 210 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Myeloid white cell count | Astle W | 27863252 | European | 2016 | 2.00E-13 | . |
| 211 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Neutrophil count | Astle W | 27863252 | European | 2016 | 4.00E-13 | . |
| 212 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Neutrophil percentage of white cells | Astle W | 27863252 | European | 2016 | 8.00E-12 | . |
| 213 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Platelet counts | Kamatani Y | 20139978 | East Asian | 2010 | 3.00E-13 | . |
| 214 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Platelet counts | Kamatani Y | 20139978 | East Asian | 2010 | 3.00E-13 | . |
| 215 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Sum basophil neutrophil counts | Astle W | 27863252 | European | 2016 | 1.00E-13 | . |
| 216 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Sum neutrophil eosinophil counts | Astle W | 27863252 | European | 2016 | 4.00E-14 | . |
| 217 | rs385893 | Yes | Strong association with blood cell traits and ambiguous whether variant is attributable to AK3 (mitochondrial gene) | *AK3;RCL1* | Platelet count | Soranzo N | 19820697 | European | 2009 | 9.00E-17 | . |
| 218 | rs1613662 | Yes | Strong associations with blood cell traits (GP6 is not a mitochondrial gene) | *GP6* | Mean platelet volume | Astle W | 27863252 | European | 2016 | 9.16E-28 | 173480 |
| 219 | rs1613662 | Yes | Strong associations with blood cell traits (GP6 is not a mitochondrial gene) | *GP6* | Platelet distribution width | Astle W | 27863252 | European | 2016 | 8.50E-45 | 173480 |
| 220 | rs1613662 | Yes | Strong associations with blood cell traits (GP6 is not a mitochondrial gene) | *GP6* | Reticulocyte count | Astle W | 27863252 | European | 2016 | 3.06E-10 | 173480 |
| 221 | rs1613662 | Yes | Strong associations with blood cell traits (GP6 is not a mitochondrial gene) | *GP6* | Reticulocyte fraction of red cells | Astle W | 27863252 | European | 2016 | 2.74E-09 | 173480 |
| 222 | rs1613662 | Yes | Strong associations with blood cell traits (GP6 is not a mitochondrial gene) | *GP6* | Mean platelet volume | Astle W | 27863252 | European | 2016 | 9.00E-28 | . |
| 223 | rs1613662 | Yes | Strong associations with blood cell traits (GP6 is not a mitochondrial gene) | *GP6* | Platelet distribution width | Astle W | 27863252 | European | 2016 | 8.00E-45 | . |

**S2. Table 9**. Mendelian Randomization analyses of mtDNA-CN versus mitochondrial disease phenotypes

| MR analysis parameters | | | | | | | MR results | | | | | | | Sensitivity Analyses | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNPs included | R2 mtDNA-CN explained by SNPs | F-statistic | Outcome Trait | FinnGen Phenocode | R2 Explained by Outcome Trait | MR Method | beta | se | OR | 95% LCI | 95% UCI | P-value | MR-PRESSO Global P-value | MR-PRESSO SNP Outliers | Q | Q df | Q P-value | intercept | intercept se | intercept P-value | correct direction? | steiger directionality p-value | |
| 1_12043717_G_A; 1_156466699_G_A; 10_102757098_G_C; 10_102761801_C_T; 10_104391034_C_T; 10_60145079_A_G; 12_109490296_G_T; 12_27867727_G_A; 14_20938251_A_C; 17_18167397_G_T; 17_18253839_C_T; 17_62476451_C_G; 19_17445208_C_T; 19_5711930_C_T; 2_74154975_G_T; 2_74166053_G_A; 2_74177777_A_G; 20_17955153_G_A; 20_17968871_C_T; 22_43528240_G_A; 22_50962208_T_G; 3_17915647_C_T; 3_48723302_G_C; 6_33546498_C_T; 7_45148667_G_A; 8_27257787_T_G; x_118601473_T_C | 0.007045374 | 100.767 | Sensorineural Hearing Loss | H8_HL_SEN_NAS | 0.0007 | **Weighted median** | **0.08** | **0.19** | **1.09** | **0.75** | **1.57** | **0.65** | 0.04 | NA | NA | NA | NA | NA | NA | NA | TRUE | 1.13E-90 | |
| | | | | | | MR Egger | -0.02 | 0.29 | 0.98 | 0.55 | 1.73 | 0.94 | | | 39.94 | 25.00 | 0.03 | 0.00 | 0.01 | 0.85 | TRUE | 1.13E-90 | |
| | | | | | | Inverse variance weighted | -0.07 | 0.16 | 0.93 | 0.69 | 1.26 | 0.66 | | | 40.00 | 26.00 | 0.04 | NA | NA | NA | TRUE | 1.13E-90 | |
| | | | Cerebrovascular Disease | FG_CEREBVASC | 0.0007 | Weighted median | 0.22 | 0.23 | 1.25 | 0.79 | 1.97 | 0.35 | 0.31 | NA | NA | NA | NA | NA | NA | NA | TRUE | 9.43E-65 | |
| | | | | | | **Inverse variance weighted** | **0.12** | **0.17** | **1.13** | **0.82** | **1.56** | **0.46** | | | 27.75 | 26.00 | 0.37 | NA | NA | NA | TRUE | 9.43E-65 | |
| | | | | | | MR Egger | -0.19 | 0.30 | 0.83 | 0.46 | 1.50 | 0.54 | | | 26.17 | 25.00 | 0.40 | -0.01 | 0.01 | 0.23 | TRUE | 9.43E-65 | |
| | | | Epilepsy | G6_EPLEPSY | 0.0009 | Weighted median | 0.16 | 0.28 | 1.17 | 0.68 | 2.03 | 0.57 | 0.52 | NA | NA | NA | NA | NA | NA | NA | TRUE | 1.18E-69 | |
| | | | | | | **Inverse variance weighted** | **0.13** | **0.20** | **1.14** | **0.78** | **1.68** | **0.49** | | | 25.40 | 26.00 | 0.50 | NA | NA | NA | TRUE | 1.18E-69 | |
| | | | | | | MR Egger | 0.05 | 0.36 | 1.05 | 0.52 | 2.13 | 0.89 | | | 25.32 | 25.00 | 0.44 | 0.00 | 0.01 | 0.78 | TRUE | 1.18E-69 | |
| | | | Migraine | G6_MIGRAINE | 0.0009 | MR Egger | -0.10 | 0.35 | 0.91 | 0.46 | 1.80 | 0.78 | 0.21 | NA | 31.48 | 25.00 | 0.17 | 0.00 | 0.01 | 0.89 | TRUE | 4.26E-73 | |
| | | | | | | **Inverse variance weighted** | **-0.14** | **0.19** | **0.87** | **0.60** | **1.26** | **0.46** | | | 31.50 | 26.00 | 0.21 | NA | NA | NA | TRUE | 4.26E-73 | |
| | | | | | | Weighted median | -0.32 | 0.26 | 0.73 | 0.44 | 1.21 | 0.22 | | | NA | NA | NA | NA | NA | NA | TRUE | 4.26E-73 | |
| | | | Dementia | F5_DEMENTIA | 0.0013 | Weighted median | 0.90 | 0.27 | 2.47 | 1.45 | 4.20 | 9.02E-04 | 0.51 | NA | NA | NA | NA | NA | NA | NA | TRUE | 1.87E-62 | |
| | | | | | | MR Egger | 0.88 | 0.36 | 2.41 | 1.19 | 4.87 | 0.02 | | | 24.67 | 25.00 | 0.48 | 0.01 | 0.01 | 0.47 | TRUE | 1.87E-62 | |
| | | | | | | **Inverse variance weighted** | **0.66** | **0.20** | **1.94** | **1.32** | **2.85** | **7.53E-04** | | | 25.20 | 26.00 | 0.51 | NA | NA | NA | TRUE | 1.87E-62 | |
| | | | Mood Disorder | F5_MOOD | 0.0005 | MR Egger | 0.17 | 0.25 | 1.19 | 0.72 | 1.95 | 0.50 | 0.03 | NA | 40.96 | 25.00 | 0.02 | 0.01 | 0.01 | 0.38 | TRUE | 4.06E-104 | |
| | | | | | | Inverse variance weighted | -0.02 | 0.14 | 0.98 | 0.75 | 1.28 | 0.89 | | | 42.28 | 26.00 | 0.02 | NA | NA | NA | TRUE | 4.06E-104 | |
| | | | | | | **Weighted median** | **-0.07** | **0.16** | **0.93** | **0.68** | **1.28** | **0.67** | | | NA | NA | NA | NA | NA | NA | TRUE | 4.06E-104 | |
| | | | Cardiomyopathy | I9_CARDMYO | 0.0015 | **Inverse variance weighted** | **0.22** | **0.28** | **1.25** | **0.72** | **2.18** | **0.43** | 0.90 | NA | 18.11 | 26.00 | 0.87 | NA | NA | NA | TRUE | 6.99E-45 | |
| | | | | | | MR Egger | -0.12 | 0.52 | 0.88 | 0.32 | 2.44 | 0.81 | | | 17.48 | 25.00 | 0.86 | -0.01 | 0.01 | 0.43 | TRUE | 6.99E-45 | |
| | | | | | | Weighted median | -0.14 | 0.38 | 0.87 | 0.42 | 1.82 | 0.71 | | | NA | NA | NA | NA | NA | NA | TRUE | 6.99E-45 | |
| | | | Conduction Disorders | I9_CONDUCTIO | 0.0016 | MR Egger | 0.97 | 0.45 | 2.64 | 1.09 | 6.38 | 0.04 | 0.32 | NA | 27.20 | 25.00 | 0.35 | 0.02 | 0.01 | 0.18 | TRUE | 2.10E-43 | |
| | | | | | | Weighted median | 0.73 | 0.33 | 2.08 | 1.10 | 3.93 | 0.02 | | | NA | NA | NA | NA | NA | NA | TRUE | 2.10E-43 | |
| | | | | | | **Inverse variance weighted** | **0.45** | **0.25** | **1.56** | **0.96** | **2.55** | **0.07** | | | 29.28 | 26.00 | 0.30 | NA | NA | NA | TRUE | 2.10E-43 | |
| | | | Type 2 diabetes, definitions combined | T2D | 0.0005 | Inverse variance weighted | 0.19 | 0.13 | 1.21 | 0.93 | 1.57 | 0.16 | 0.04 | 10_102761801_c_t | 40.32 | 26.00 | 0.04 | NA | NA | NA | TRUE | 5.06E-103 | |
| | | | | | | **Weighted median** | **0.14** | **0.16** | **1.15** | **0.84** | **1.57** | **0.37** | | | NA | NA | NA | NA | NA | NA | TRUE | 5.06E-103 | |
| | | | | | | MR Egger | -0.13 | 0.24 | 0.87 | 0.55 | 1.40 | 0.58 | | | 36.54 | 25.00 | 0.06 | -0.01 | 0.01 | 0.12 | TRUE | 5.06E-103 | |
| | | | Paralytic ileus and intestinal obstruction | K11_ILEUS | 0.0020 | Weighted median | 0.03 | 0.40 | 1.04 | 0.47 | 2.28 | 0.93 | 0.08 | NA | NA | NA | NA | NA | NA | NA | TRUE | 1.73E-38 | |
| | | | | | | **Inverse variance weighted** | **-0.47** | **0.28** | **0.63** | **0.36** | **1.07** | **0.09** | | | 34.04 | 26.00 | 0.13 | NA | NA | NA | TRUE | 1.73E-38 | |
| | | | | | | MR Egger | -0.54 | 0.52 | 0.58 | 0.21 | 1.59 | 0.30 | | | 34.00 | 25.00 | 0.11 | 0.00 | 0.01 | 0.86 | TRUE | 1.73E-38 | |

## Supplementary Figures

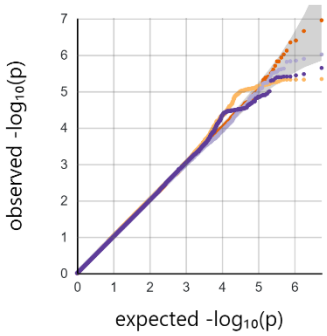
**African** **South Asian** **Other White**




0.01 ≤ MAF < 0.02 (4745474)
0.02 ≤ MAF < 0.05 (4745474)
0.05 ≤ MAF < 0.16 (4745474)
0.16 ≤ MAF < 0.50 (4745474)

**GC lambda 0.5: 1.010**
GC lambda 0.1: 1.011
GC lambda 0.01: 1.008
GC lambda 0.001: 1.010
*(Genomic Control lambda calculated based on the 50th percentile (median), 10th percentile, 1st percentile, and 1/10th of a percentile)*

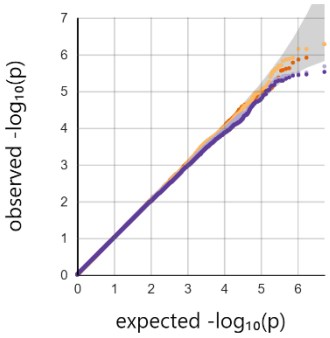
**Irish**




0.01 ≤ MAF < 0.02 (2837745)
0.02 ≤ MAF < 0.09 (2837745)
0.09 ≤ MAF < 0.26 (2837745)
0.26 ≤ MAF < 0.50 (2837746)

**GC lambda 0.5: 1.008**
GC lambda 0.1: 1.009
GC lambda 0.01: 1.003
GC lambda 0.001: 1.002
*(Genomic Control lambda calculated based on the 50th percentile (median), 10th percentile, 1st percentile, and 1/10th of a percentile)*

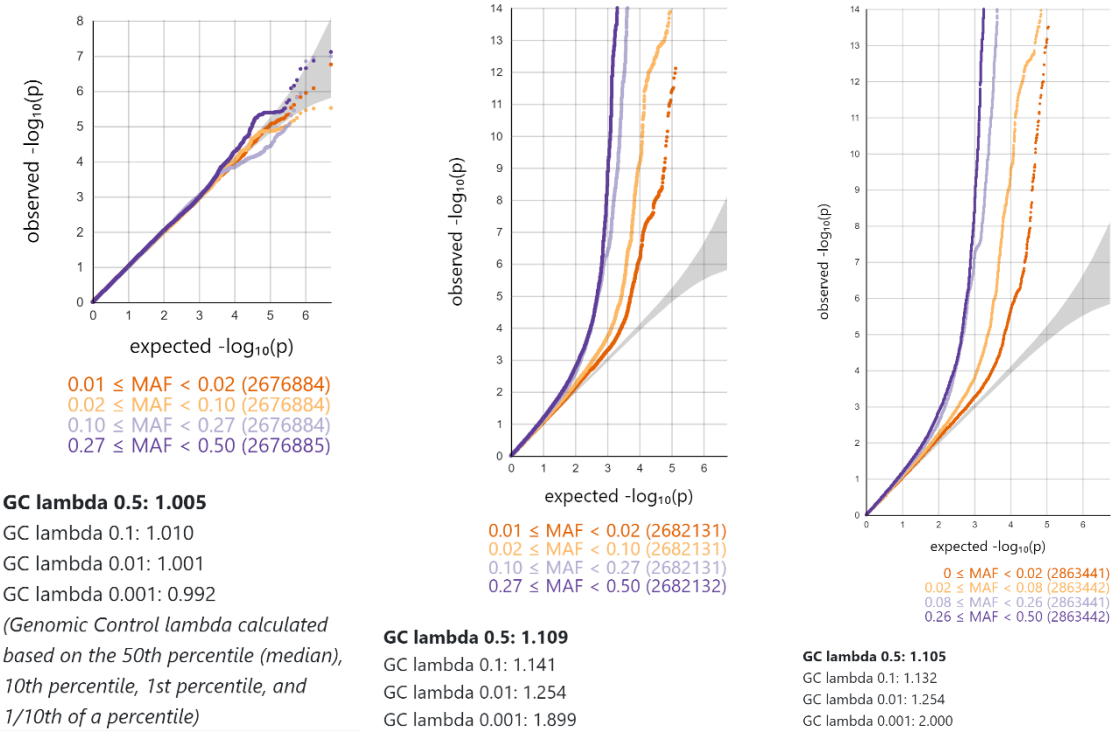
**British**




0.01 ≤ MAF < 0.02 (2723624)
0.02 ≤ MAF < 0.10 (2723624)
0.10 ≤ MAF < 0.26 (2723624)
0.26 ≤ MAF < 0.50 (2723625)

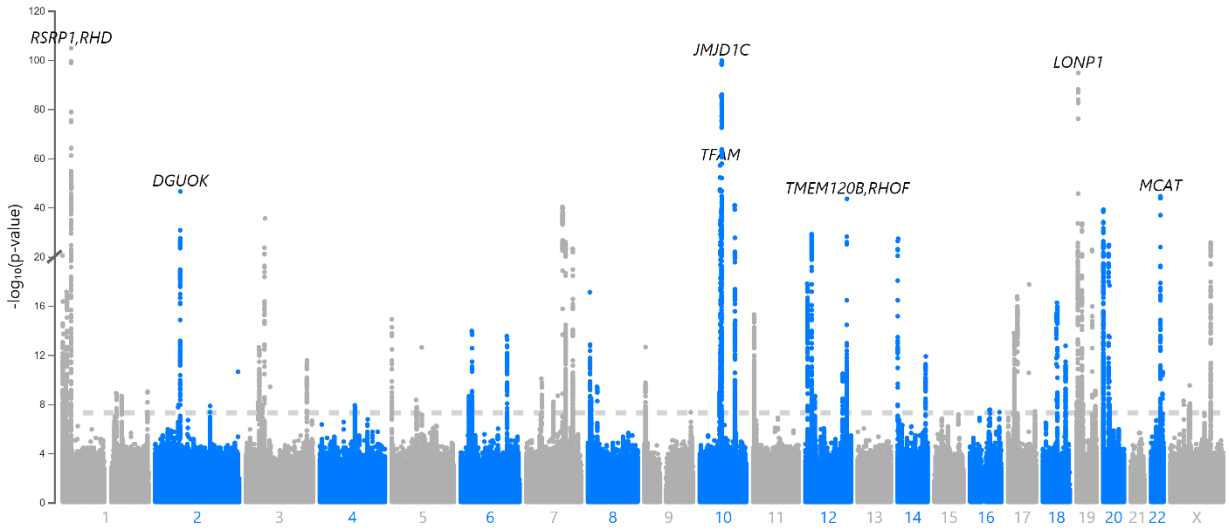
**GC lambda 0.5: 1.009**
GC lambda 0.1: 1.010
GC lambda 0.01: 1.008
GC lambda 0.001: 0.999
*(Genomic Control lambda calculated based on the 50th percentile (median), 10th percentile, 1st percentile, and 1/10th of a percentile)*

**European Meta-Analysis**

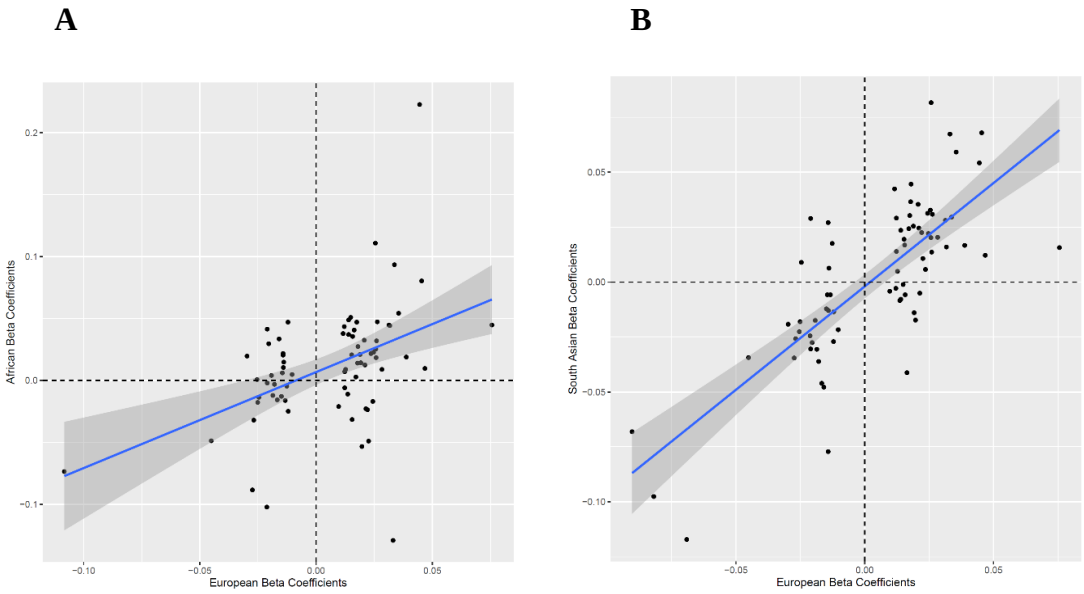



**S2. Figure 1**. MAF and ethnicity-stratified GWAS quantile-quantile plots.

**S2. Figure 2 (Extended Figures)**. Locus zoom plots for 72 loci and 82 conditionally independent genetic signals. Variants with the highest fine-mapping posterior probability are labelled by their genomic coordinates (GRCh37). Pairwise correlation between the lead variant and proximal variants were colour-coded based on the 1000Genomes Europeans reference panel.

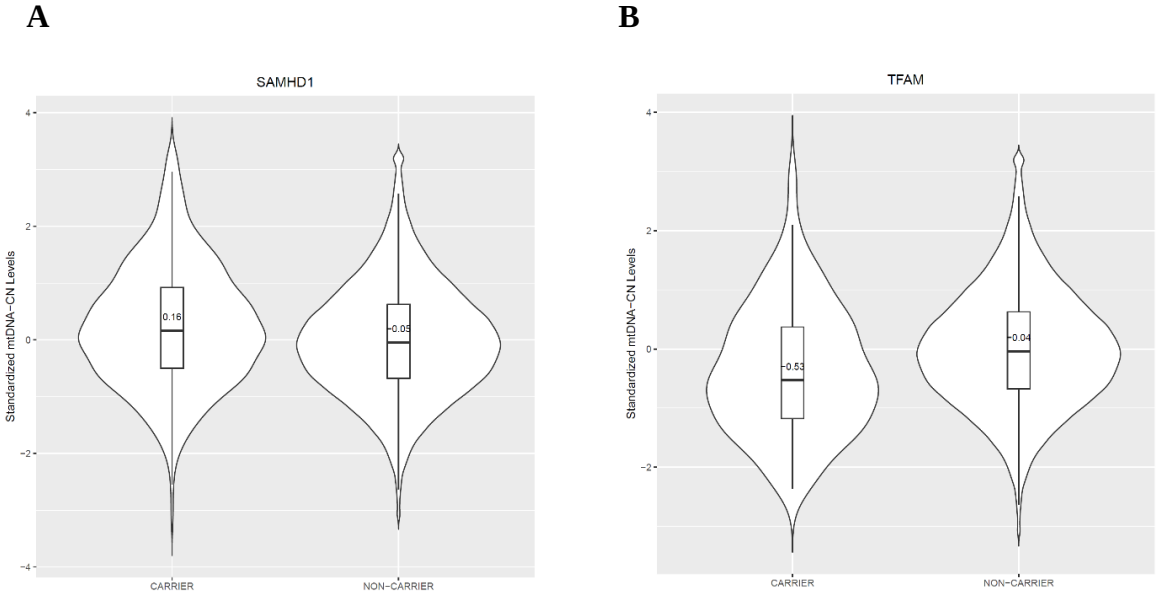



**S2 Figure 3.** Manhattan plot for trans-ethnic GWAS meta-analysis (N=395,781).




**S2 Figure 4.** Correlation between conditionally independent mtDNA-CN loci effect estimates derived from European GWAS meta-analyses (x-axes) vs. effect estimates from Non-European GWAS (y-axes). Comparisons for African (A) and South Asian (B) GWAS analyses are presented. Of the total 82 conditionally independent signals identified using

the European GWAS meta-analysis, 73 and 75 variants were available for comparison in African and South Asian GWAS, respectively.




**S2. Figure 5**. Violin plots showing the distribution of mtDNA-CN for carriers and non-carriers of (A) *SAMHD1* and (B) *TFAM* rare nonsynonymous and deleterious (MCAP > 0.025) variants.

## Supplementary References

Abecasis, Goncalo R et al. 2012. "An Integrated Map of Genetic Variation from 1,092 Human Genomes." *Nature* 491(7422): 56–65. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3498066&tool=pmcentrez&rendertype=abstract (July 9, 2014).

Ali, Aminah T. et al. 2019. "Nuclear Genetic Regulation of the Human Mitochondrial Transcriptome." *eLife* 8: 1–23.

Altschul, S F et al. 1990. "Basic Local Alignment Search Tool." *Journal of molecular*

*biology* 215(3): 403–10. http://www.ncbi.nlm.nih.gov/pubmed/2231712.

Ashar, Foram N et al. 2017. "Association of Mitochondrial DNA Copy Number With Cardiovascular Disease." 21205(11): 1247–55.

Benner, Christian et al. 2016. "FINEMAP: Efficient Variable Selection Using Summary Data from Genome-Wide Association Studies." *Bioinformatics* 32(10): 1493–1501.

———. 2017. "Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-Wide Association Studies." *The American Journal of Human Genetics* 101(4): 539–51. https://linkinghub.elsevier.com/retrieve/pii/S0002929717303348.

Bulik-Sullivan, Brendan et al. 2015. "LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies." *Nature Genetics* 47(3): 291–95.

Bycroft, Clare et al. 2018. "The UK Biobank Resource with Deep Phenotyping and Genomic Data." *Nature* 562(7726): 203–9. http://www.nature.com/articles/s41586-018-0579-z.

Denny, Joshua C et al. 2013. "Systematic Comparison of Phenome-Wide Association Study of Electronic Medical Record Data and Genome-Wide Association Study Data." *Nature biotechnology* 31(12): 1102–10. http://www.ncbi.nlm.nih.gov/pubmed/24270849.

Diskin, Sharon J et al. 2008. "Adjustment of Genomic Waves in Signal Intensities from Whole-Genome SNP Genotyping Platforms." *Nucleic acids research* 36(19): e126. http://www.ncbi.nlm.nih.gov/pubmed/18784189.

Fazzini, Federica et al. 2019. “Mitochondrial DNA Copy Number Is Associated with Mortality and Infections in a Large Cohort of Patients with Chronic Kidney Disease.” 8: 480–88.

Feng, Yen-Chen A et al. 2020. “Findings and Insights from the Genetic Investigation of Age of First Reported Occurrence for Complex Disorders in the UK Biobank and FinnGen.” *medRxiv*: 2020.11.20.20234302. https://doi.org/10.1101/2020.11.20.20234302.

Gene, The, and Ontology Consortium. 2000. “Gene Ontology : Tool for The.” 25(may): 25–29.

Gorman, Gráinne S. et al. 2016. “Mitochondrial Diseases.” *Nature Reviews Disease Primers* 2.

GTEX. 2014. “The Genotype-Tissue Expression (GTEx) Project The.” *Nature Genetics* 45(6): 580–85.

Harris, Charles R. et al. 2020. “Array Programming with NumPy.” *Nature* 585(7825): 357–62. http://www.nature.com/articles/s41586-020-2649-2.

Hemani, Gibran et al. 2018. “The MR-Base Platform Supports Systematic Causal Inference across the Human Phenome.” *eLife* 7: 1–29.

Jagadeesh, Karthik A et al. 2016. “M-CAP Eliminates a Majority of Variants of Uncertain Significance in Clinical Exomes at High Sensitivity.” *Nature genetics* 48(12): 1581–86. http://www.ncbi.nlm.nih.gov/pubmed/27776117.

Kamat, Mihir A et al. 2019. “PhenoScanner V2: An Expanded Tool for Searching Human Genotype-Phenotype Associations.” *Bioinformatics (Oxford, England)* 35(22):

4851–53. http://www.ncbi.nlm.nih.gov/pubmed/31233103.

Lambert, J C et al. 2013. “Meta-Analysis of 74,046 Individuals Identifies 11 New Susceptibility Loci for Alzheimer’s Disease.” *Nature genetics* 45(12): 1452–58. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3896259&tool=pmcentrez&rendertype=abstract (July 14, 2014).

Landrum, M. J. et al. 2015. “ClinVar: Public Archive of Interpretations of Clinically Relevant Variants.” *Nucleic Acids Research*. http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkv1222.

Lane, John. 2014. “MitoPipeline: Generating Mitochondrial Copy Number Estimates from SNP Array Data in Genvisis.” http://genvisis.org/MitoPipeline/.

Longchamps, Ryan Joseph. 2019. “EXPLORING THE ROLE OF MITOCHONDRIAL DNA QUANTITY AND QUALITY.” *Biorxiv* (August).

Mbatchou, Joelle et al. 2020. “Computationally Efficient Whole Genome Regression for Quantitative and Binary Traits.” : 1–88.

O’Donnell, Martin J et al. 2010. “Risk Factors for Ischaemic and Intracerebral Haemorrhagic Stroke in 22 Countries (the INTERSTROKE Study): A Case-Control Study.” *Lancet* 376(9735): 112–23. http://www.ncbi.nlm.nih.gov/pubmed/20561675 (July 11, 2014).

Pers, Tune H. et al. 2015. “Biological Interpretation of Genome-Wide Association Studies Using Predicted Gene Functions.” *Nature Communications* 6: 5890. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4420238&tool=pmcentrez&rendertype=abstract.

Pruim, Randall J et al. 2010. “LocusZoom : Regional Visualization of Genome-Wide Association Scan Results.” 26(18): 2336–37.

Purcell, Shaun et al. 2007. “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.” *American journal of human genetics* 81(3): 559–75.

Rath, Sneha et al. 2021. “MitoCarta3.0: An Updated Mitochondrial Proteome Now with Sub-Organelle Localization and Pathway Annotations.” *Nucleic acids research* 49(D1): D1541–47.

Rentzsch, Philipp et al. 2019. “CADD: Predicting the Deleteriousness of Variants throughout the Human Genome.” *Nucleic Acids Research* 47(D1): D886–94. https://academic.oup.com/nar/article/47/D1/D886/5146191.

Simone, Domenico et al. 2011. “The Reference Human Nuclear Mitochondrial Sequences Compilation Validated and Implemented on the UCSC Genome Browser.” *BMC genomics* 12: 517. http://www.ncbi.nlm.nih.gov/pubmed/22013967.

Sudlow, Cathie et al. 2015. “UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age.” *PLoS medicine* 12(3): e1001779. http://www.ncbi.nlm.nih.gov/pubmed/25826379.

Verbanck, Marie, Chia Yen Chen, Benjamin Neale, and Ron Do. 2018. “Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred from Mendelian Randomization between Complex Traits and Diseases.” *Nature Genetics* 50(5): 693–98. http://dx.doi.org/10.1038/s41588-018-0099-7.

Virtanen, Pauli et al. 2020. “SciPy 1.0: Fundamental Algorithms for Scientific Computing

in Python.” *Nature Methods* 17(3): 261–72. http://www.nature.com/articles/s41592-019-0686-2.

Vuckovic, Dragana et al. 2020. “The Polygenic and Monogenic Basis of Blood Traits and Diseases.” *Cell* 182(5): 1214-1231.e11. http://www.ncbi.nlm.nih.gov/pubmed/32888494.

Wang, Kai et al. 2007. “PennCNV: An Integrated Hidden Markov Model Designed for High-Resolution Copy Number Variation Detection in Whole-Genome SNP Genotyping Data.” *Genome research* 17(11): 1665–74. http://www.ncbi.nlm.nih.gov/pubmed/17921354.

Warde-Farley, David et al. 2010. “The GeneMANIA Prediction Server: Biological Network Integration for Gene Prioritization and Predicting Gene Function.” *Nucleic Acids Research* 38(SUPPL. 2): 214–20.

Wei, Wei Qi et al. 2017. “Evaluating Phecodes, Clinical Classification Software, and ICD-9-CM Codes for Phenome-Wide Association Studies in the Electronic Health Record.” *PLoS ONE* 12(7): 1–16.

Willer, Cristen J., Yun Li, and Gonçalo R. Abecasis. 2010. “METAL: Fast and Efficient Meta-Analysis of Genomewide Association Scans.” *Bioinformatics* 26(17): 2190–91.

Wu, Patrick et al. 2019. “Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation.” *JMIR medical informatics* 7(4): e14325. http://www.ncbi.nlm.nih.gov/pubmed/31553307.

Zhang, Yiyi et al. 2017. “Association between Mitochondrial DNA Copy Number and

Sudden Cardiac Death : Findings from the Atherosclerosis Risk in Communities Study ( ARIC ).” : 3443–48.

Zhou, Wei et al. 2018. “Efficiently Controlling for Case-Control Imbalance and Sample Relatedness in Large-Scale Genetic Association Studies.” *Nature Genetics* 50(9): 1335–41. http://www.nature.com/articles/s41588-018-0184-y.

# APPENDIX C:

## Supplementary Data for Study 3

### Supplementary Methods

***mtDNA-CN measurement in INTERSTROKE***

mtDNA-CN was assayed by the Genetic and Molecular Epidemiology Lab located in Hamilton, Ontario, Canada using a plasmid-normalized quantitative Polymerase Chain Reaction (qPCR) method developed by Fazzini *et al.* The qPCR assay is a duplex assay that simultaneously amplifies segments of the mitochondrial *tRNALeu* and the nuclear *B2M* genes. A plasmid construct containing a single copy of both mitochondrial and nuclear fragments was amplified to calibrate differences in fluorescent dye intensities used to differentiate mitochondrial and nuclear signals. All INTERSTROKE study participants' samples were assayed in duplicate, and any samples demonstrating inconsistent cycle thresholds with coefficient of variation greater than 5% were excluded. Four calibrator samples were included on each of the qPCR plates to measure assay reproducibility. The mean coefficient of variation among these four samples across plates was 6.5%, which is consistent with 7% intra-assay variability previously reported for this protocol[1].

***Mendelian Randomization Instrument Selection***

Selection of genetic variants was based on a previously described scheme[2]. First, to satisfy the MR "relevance" assumption, we extracted 9602 genome-wide significant ($P < 5x10^{-8}$) variants associated with buffy coat mtDNA-CN levels from the UKBiobank GWAS. Second, in consideration of the "exclusion restriction" assumption, we removed variants located outside of genes (> 100kb from the transcript) encoding for proteins localizing to the mitochondria to exclude variants that may act through non-mitochondrial pathways, thus resulting in 1701 variants [3]. Third, we matched these variants to the GISCOME outcome GWAS and retained 1416 shared variants between UKBiobank and

GISCOME. Fourth, we performed linkage disequilibrium pruning (1000Genomes Europeans LD r2 < 0.01) to achieve an independent set of 33 variants. Fifth, to further mitigate potential for horizontal pleiotropy, we performed a phenome-wide search across published GWAS with Phenoscanner V2 [4] and removed seven variants with strong evidence of acting through alternative pathways ($P<5x10^{-20}$) [3]. Ultimately, 26 genetic variants remained as suitable genetic instruments.

***Genetic instruments for neutrophil to lymphocyte ratio***

Whole genome model linear regressions were performed using the REGENIE program in 340002 British participants from the UKBiobank study (unpublished data)[5]. Genotyped and imputed variants with minor allele frequency > 0.001 and quality INFO score > 0.3 were analyzed. Association analyses were adjusted for the following covariates: age, age$^2$, sex, 20 ancestry-informative principal components, type of genotyping array and assessment centre. Like the other MR analyses, an independent (1000Genomes European LD r2 < 0.01) set of genome-wide significant ($P<5x10^{-8}$) variants were retained to approximate genetically determined neutrophil to lymphocyte ratio. Causal effects on mRS-based outcomes are expressed in terms of a 1 SD increase in the neutrophil to lymphocyte ratio.

## Supplementary Table

**S. Table 1.** Association between mtDNA-CN and baseline stroke severity characteristics. Analyses of mtDNA-CN as a continuous variable are expressed per 1 SD decrease in mtDNA-CN. mtDNA-CN quartile comparisons are expressed with reference to the highest ($4^{th}$) quartile.

| Analysis Parameters | | | | Main Summary Statistics | | | | | | Proportional Odds Check |
|---|---|---|---|---|---|---|---|---|---|---|
| *mtDNA-CN Quartile (or Continuous)* | *Baseline Stroke Characteristic* | $N_{no\ outcome}$ | $N_{outcome}$ | *Beta* | *SE* | *OR* | *LCI* | *UCI* | *P-value* | *Brant test P-value* |
| Continuous | Ordinal mRS | NA | 3493 | 0.24 | 0.03 | 1.27 | 1.19 | 1.36 | 4.67E-12 | 0.77 |
| Continuous | Reduced Consciousness | NA | 3491 | 0.29 | 0.05 | 1.33 | 1.21 | 1.48 | 1.76E-08 | 0.97 |
| Continuous | Hemorrhagic Transformation | 2845 | 54 | 0.29 | 0.19 | 1.33 | 0.92 | 1.93 | 0.13 | NA |
| 1st | Ordinal mRS | NA | 874 | 0.70 | 0.10 | 2.00 | 1.65 | 2.44 | 2.86E-12 | 0.53 |
| 2nd | Ordinal mRS | NA | 873 | 0.50 | 0.10 | 1.66 | 1.37 | 2.01 | 2.48E-07 | 0.91 |
| 3rd | Ordinal mRS | NA | 872 | 0.30 | 0.10 | 1.34 | 1.11 | 1.62 | 1.98E-03 | 0.37 |
| 4th | Ordinal mRS | NA | 874 | NA | NA | NA | NA | NA | NA | NA |
| 1st | Reduced Consciousness | NA | 873 | 0.88 | 0.14 | 2.42 | 1.84 | 3.17 | 1.58E-10 | 0.31 |
| 2nd | Reduced Consciousness | NA | 873 | 0.69 | 0.14 | 1.99 | 1.50 | 2.64 | 1.76E-06 | 0.35 |
| 3rd | Reduced Consciousness | NA | 873 | 0.33 | 0.15 | 1.39 | 1.04 | 1.87 | 2.67E-02 | 0.37 |
| 4th | Reduced Consciousness | NA | 872 | NA | NA | NA | NA | NA | NA | NA |
| 1st | Hemorrhagic Transformation | 642 | 20 | 0.82 | 0.46 | 2.28 | 0.92 | 5.66 | 0.08 | NA |
| 2nd | Hemorrhagic Transformation | 696 | 13 | 0.06 | 0.51 | 1.06 | 0.39 | 2.90 | 0.91 | NA |
| 3rd | Hemorrhagic Transformation | 749 | 11 | 0.07 | 0.53 | 1.07 | 0.38 | 3.03 | 0.90 | NA |
| 4th | Hemorrhagic Transformation | 758 | 10 | NA | NA | NA | NA | NA | NA | NA |

**S. Table 2.** Comparison of the association between mtDNA-CN (per 1 SD decrease) and 1-month post-stroke outcomes versus age (per 1 SD increase) and 1-month post-stroke outcomes.

| Analysis Parameters | | | | Main Summary Statistics | | | | | | Proportional Odds Check |
|---|---|---|---|---|---|---|---|---|---|---|
| *Exposure* | *1-month Stroke Outcome* | $N_{w/\ outcome}$ | $N_{w/o\ outcome}$ | *Beta* | *SE* | *OR* | *LCI* | *UCI* | *P-value* | *Brant test P-value* |
| mtDNA-CN* | Ordinal mRS | 3470 | NA | 0.15 | 0.04 | 1.16 | 1.08 | 1.24 | 4.41E-05 | 0.2 |
| Age** | Ordinal mRS | 3470 | NA | 0.25 | 0.04 | 1.28 | 1.18 | 1.39 | 1.77E-09 | 0.11 |
| mtDNA-CN* | Poor Functional Outcome | 1354 | 2116 | 0.19 | 0.06 | 1.21 | 1.08 | 1.34 | 6.85E-04 | NA |
| Age** | Poor Functional Outcome | 1354 | 2116 | 0.34 | 0.06 | 1.41 | 1.25 | 1.59 | 4.10E-08 | NA |
| mtDNA-CN* | Mortality | 337 | 3133 | 0.30 | 0.08 | 1.35 | 1.14 | 1.59 | 3.90E-04 | NA |
| Age** | Mortality | 337 | 3133 | 0.16 | 0.08 | 1.18 | 1.00 | 1.38 | 4.62E-02 | NA |

*Effects are expressed per 1 SD decrease in mtDNA-CN
**Effects are expressed per 1SD increase in age

**S. Table 3.** mtDNA-CN quartile associations with 1-month post-stroke outcomes. Results are expressed with reference to the highest ($4^{th}$) quartile.

| Analysis Parameters | | | | Main Summary Statistics | | | | | | Proportional Odds Check |
|---|---|---|---|---|---|---|---|---|---|---|
| *mtDNA-CN Quartile* | *1-month Stroke Outcome* | $N_{w/\ outcome}$ | $N_{w/o\ outcome}$ | *Beta* | *SE* | *OR* | *LCI* | *UCI* | *P-value* | *Brant test P-value* |
| 1st | Ordinal mRS | 869 | NA | 0.34 | 0.10 | 1.40 | 1.15 | 1.71 | 9.02E-04 | 0.49 |
| 2nd | Ordinal mRS | 865 | NA | 0.29 | 0.10 | 1.34 | 1.10 | 1.63 | 3.35E-03 | 0.41 |
| 3rd | Ordinal mRS | 866 | NA | 0.00 | 0.10 | 1.00 | 0.83 | 1.22 | 9.88E-01 | 0.25 |
| 4th | Ordinal mRS | 869 | NA | NA | NA | NA | NA | NA | NA | NA |
| 1st | Poor Functional Outcome | 452 | 417 | 0.41 | 0.16 | 1.51 | 1.11 | 2.04 | 8.34E-03 | NA |
| 2nd | Poor Functional Outcome | 368 | 497 | 0.27 | 0.15 | 1.30 | 0.97 | 1.76 | 8.38E-02 | NA |
| 3rd | Poor Functional Outcome | 284 | 582 | -0.09 | 0.15 | 0.92 | 0.68 | 1.24 | 5.76E-01 | NA |
| 4th | Poor Functional Outcome | 250 | 619 | NA | NA | NA | NA | NA | NA | NA |
| 1st | Mortality | 146 | 723 | 0.74 | 0.23 | 2.09 | 1.34 | 3.25 | 1.10E-03 | NA |
| 2nd | Mortality | 95 | 770 | 0.59 | 0.24 | 1.80 | 1.12 | 2.89 | 1.51E-02 | NA |
| 3rd | Mortality | 52 | 814 | 0.07 | 0.27 | 1.07 | 0.64 | 1.81 | 7.87E-01 | NA |
| 4th | Mortality | 44 | 825 | NA | NA | NA | NA | NA | NA | NA |

**S. Table 4.** Subgroup analyses for mtDNA-CN associations with 1-month post-stroke outcomes. Poor functional outcome is defined as 1-month mRS 3-6. Odds ratios are expressed per 1 SD decrease in mtDNA-CN.

| Strata | | | | | Main Summary Statistics | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Variable* | *Subgroup* | *1-month Stroke Outcome* | $N_{outcome}$ | $N_{non\text{-}outcome}$ | *Beta* | *SE* | *OR* | *LCI* | *UCI* | *P-value* | $I^2$ | *Heterogeneity P-value* |
| Primary Stroke Type | Ischemic Stroke | Poor Functional Outcome | 916 | 1701 | 0.18 | 0.06 | 1.20 | 1.07 | 1.36 | 0.00262 | 0 | |
| Primary Stroke Type | Hemorrhagic Stroke | Poor Functional Outcome | 303 | 245 | 0.18 | 0.14 | 1.19 | 0.91 | 1.57 | 0.20440 | | 0.9612 |
| Sex | Female | Poor Functional Outcome | 638 | 828 | 0.22 | 0.08 | 1.25 | 1.06 | 1.46 | 0.00615 | | |
| Sex | Male | Poor Functional Outcome | 712 | 1288 | 0.18 | 0.08 | 1.19 | 1.02 | 1.39 | 0.02466 | 0 | 0.6926 |
| Age | <= 65 years | Poor Functional Outcome | 549 | 1183 | 0.17 | 0.08 | 1.18 | 1.00 | 1.39 | 0.04384 | | |
| Age | > 65 years | Poor Functional Outcome | 801 | 933 | 0.20 | 0.08 | 1.23 | 1.06 | 1.42 | 0.00695 | 0 | 0.7418 |
| World Bank Country Income | Lower-middle/Low | Poor Functional Outcome | 326 | 339 | 0.17 | 0.09 | 1.19 | 1.00 | 1.41 | 0.04919 | 10.71 | 0.3263 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| World Bank Country Income | Upper-middle | Poor Functional Outcome | 633 | 615 | 0.28 | 0.10 | 1.32 | 1.09 | 1.61 | 0.00526 | | |
| World Bank Country Income | High | Poor Functional Outcome | 391 | 1162 | 0.06 | 0.11 | 1.06 | 0.86 | 1.31 | 0.59446 | | |
| Education Level | None or Primary | Poor Functional Outcome | 531 | 572 | 0.17 | 0.09 | 1.18 | 0.99 | 1.41 | 0.06105 | 0 | 0.8119 |
| Education Level | High School, Trade School, College, or University | Poor Functional Outcome | 692 | 1374 | 0.20 | 0.07 | 1.22 | 1.05 | 1.40 | 0.00706 | | |
| All | All | Poor Functional Outcome | 1354 | 2116 | 0.19 | 0.06 | 1.21 | 1.08 | 1.34 | 0.00068 | NA | NA |
| Primary Stroke Type | Ischemic Stroke | Mortality | 189 | 2428 | 0.29 | 0.10 | 1.33 | 1.08 | 1.63 | 0.00641 | 0 | 0.9992 |
| Primary Stroke Type | Hemorrhagic Stroke | Mortality | 117 | 431 | 0.29 | 0.15 | 1.33 | 1.00 | 1.77 | 0.05231 | | |
| Sex | Female | Mortality | 161 | 1305 | 0.34 | 0.12 | 1.41 | 1.11 | 1.79 | 0.00468 | 0 | 0.7113 |
| Sex | Male | Mortality | 172 | 1828 | 0.28 | 0.12 | 1.32 | 1.04 | 1.68 | 0.02137 | | |
| Age | <= 65 years | Mortality | 143 | 1589 | 0.44 | 0.13 | 1.56 | 1.20 | 2.01 | 0.00071 | 57.54 | 0.1249 |
| Age | > 65 years | Mortality | 190 | 1544 | 0.18 | 0.11 | 1.20 | 0.97 | 1.49 | 0.09740 | | |
| World Bank Country Income | Lower-middle/Low | Mortality | 125 | 540 | 0.21 | 0.11 | 1.23 | 0.99 | 1.53 | 0.06581 | 0 | 0.514 |
| World Bank Country Income | Upper-middle | Mortality | 175 | 1073 | 0.37 | 0.15 | 1.45 | 1.09 | 1.93 | 0.00993 | | |
| World Bank Country Income | High | Mortality | 33 | 1520 | 0.55 | 0.39 | 1.74 | 0.81 | 3.72 | 0.15523 | | |
| Education Level | None or Primary | Mortality | 159 | 944 | 0.33 | 0.13 | 1.39 | 1.08 | 1.78 | 0.00941 | 0 | 0.7923 |
| Education Level | High School, Trade School, College, or University | Mortality | 151 | 1915 | 0.28 | 0.12 | 1.33 | 1.06 | 1.67 | 0.01368 | | |
| All | All | Mortality | 337 | 3133 | 0.30 | 0.08 | 1.35 | 1.14 | 1.59 | 0.00039 | NA | NA |

**S. Table 5.** Net reclassification indices for the addition of mtDNA-CN into logistic regression models. Model covariates include age, sex, region, education level (none or primary school vs. high school, trade school, college, or university), 2018 World Bank country income stratum (high, upper-middle, and lower-middle or low income), household income (adjusted for country), primary stroke type (ischemic vs. hemorrhagic stroke) and ischemic stroke Oxfordshire Community Stroke Project (OCSP) classification, pre-stroke dependency (pre-stroke mRS 3-5 vs. 0-2), Charleson comorbidity index, and stroke risk factors (hypertension, diabetes, hypercholesterolemia, atrial fibrillation or flutter, current smoker status, and waist to hip ratio), and baseline mRS.

| 1-month Stroke Outcome | $N_{event}$ | $N_{non\text{-}event}$ | NRI Strata | NRI | 95% LCI | 95% UCI | P-value |
|---|---|---|---|---|---|---|---|
| Poor Functional Outcome | 1219 | 1945 | $NRI_{overall}$ | 0.16 | 0.08 | 0.23 | 3.60E-05 |
| Poor Functional Outcome | 1219 | 1945 | $NRI_{events}$ | 0.20 | 0.15 | 0.26 | 4.30E-12 |
| Poor Functional Outcome | 1219 | 1945 | $NRI_{non\_events}$ | -0.05 | -0.09 | 0.00 | 4.50E-02 |
| Mortality | 306 | 2858 | $NRI_{overall}$ | 0.31 | 0.19 | 0.43 | 1.70E-07 |
| Mortality | 306 | 2858 | $NRI_{events}$ | 0.33 | 0.22 | 0.44 | 3.40E-09 |
| Mortality | 306 | 2858 | $NRI_{non\_events}$ | -0.02 | -0.06 | 0.02 | 2.96E-01 |

**S. Table 6.** Characteristics of 33 independent genetic variants located within 100kb of MitoCarta3-annotated genes considered for MR analyses. A phenoscanner v2 search was performed, identifying seven variants strongly associated with other traits ($P<5x10^{-20}$). 26 genetic variants were retained as suitable genetic instruments to approximate genetically determined mtDNA-CN levels.

| rsid | snpid | chr | pos_hg19 | ref | alt | beta | se | p-value | Nearest Gene(s) | 1000G Eur AF | Phenoscanner v2 Pleiotropy Detected? | Included in MR? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs3766744 | 1_12043717_G_A | 1 | 12043717 | G | A | -0.0179 | 0.0021 | 2.91E-17 | MFN2 | 0.4722 | No | Yes |
| rs1342442 | 1_156466699_G_A | 1 | 156466699 | G | A | -0.0138 | 0.0022 | 5.56E-10 | MEF2D | 0.675 | No | Yes |
| rs11677402 | 2_74154975_G_T | 2 | 74154975 | G | T | 0.0251 | 0.0028 | 7.37E-20 | DGUOK | 0.6332 | No | Yes |
| rs62641680 | 2_74166053_G_A | 2 | 74166053 | G | A | -0.0903 | 0.0063 | 4.63E-47 | DGUOK | 0.0249 | No | Yes |
| rs74874677 | 2_74177777_A_G | 2 | 74177777 | A | G | -0.0819 | 0.0071 | 3.62E-31 | DGUOK | 0.0298 | No | Yes |
| rs6792510 | 3_48723302_G_C | 3 | 48723302 | G | C | -0.0126 | 0.0022 | 1.44E-08 | NCKIPSD | 0.6471 | No | Yes |
| rs13088724 | 3_179152841_G_A | 3 | 179152841 | G | A | 0.0175 | 0.0026 | 7.70E-12 | GNB4 | 0.2714 | No | Yes |
| rs2844509 | 6_31510924_A_G | 6 | 31510924 | A | G | 0.014 | 0.0025 | 2.61E-08 | ATP6V1G2-DDX39B | 0.2704 | Yes | No |
| rs5745582 | 6_33546498_C_T | 6 | 33546498 | C | T | 0.021 | 0.0028 | 3.46E-14 | BAK1 | 0.2247 | No | Yes |
| rs2304693 | 7_45148667_G_A | 7 | 45148667 | G | A | 0.0181 | 0.0028 | 5.11E-11 | TBRG4; RAMP3 | 0.2097 | No | Yes |
| rs2322718 | 8_27257787_T_G | 8 | 27257787 | T | G | 0.0136 | 0.0021 | 1.53E-10 | PTK2B | 0.4533 | No | Yes |
| rs385893 | 9_4763176_T_C | 9 | 4763176 | T | C | 0.0153 | 0.0021 | 5.96E-13 | AK3; RCL1 | 0.493 | No | No |
| rs8176645 | 9_136149098_T_A | 9 | 136149098 | T | A | -0.0145 | 0.0026 | 3.06E-08 | ABO | 0.3966 | Yes | No |
| rs12247015 | 10_60145079_A_G | 10 | 60145079 | A | G | 0.0337 | 0.0021 | 1.28E-55 | TFAM | 0.4245 | No | Yes |
| rs57066921 | 10_102752345_T_G | 10 | 102752345 | T | G | -0.1081 | 0.0081 | 8.98E-41 | LZTS2 | 0.0169 | No | Yes |
| rs701834 | 10_102761801_C_T | 10 | 102761801 | C | T | -0.0254 | 0.0026 | 1.12E-22 | LZTS2 | 0.1839 | No | Yes |
| rs11596235 | 10_104391034_C_T | 10 | 104391034 | C | T | 0.0174 | 0.0023 | 2.16E-14 | SUFU | 0.3509 | No | Yes |
| rs1127787 | 12_27867727_G_A | 12 | 27867727 | G | A | -0.0159 | 0.0028 | 1.49E-08 | MRPS35 | 0.173 | No | Yes |
| rs12426673 | 12_109490296_G_T | 12 | 109490296 | G | T | -0.0141 | 0.0021 | 4.03E-11 | USP30 | 0.6233 | No | Yes |
| rs1760940 | 14_20938251_A_C | 14 | 20938251 | A | C | 0.0263 | 0.0024 | 5.20E-27 | PNP | 0.2455 | No | Yes |
| rs3889402 | 17_18167397_G_T | 17 | 18167397 | G | T | 0.026 | 0.0034 | 2.60E-14 | MIEF2 | 0.0924 | No | Yes |
| rs62072546 | 17_18253839_C_T | 17 | 18253839 | C | T | -0.0149 | 0.0025 | 2.95E-09 | SHMT1 | 0.2227 | No | Yes |
| rs8067252 | 17_29263700_C_T | 17 | 29263700 | C | T | -0.0142 | 0.0026 | 3.58E-08 | ADAP2 | 0.2276 | Yes | No |
| rs17850455 | 17_62476451_C_G | 17 | 62476451 | C | G | 0.091 | 0.0104 | 1.81E-18 | POLG2 | 0.0129 | No | Yes |
| rs11085147 | 19_5711930_C_T | 19 | 5711930 | C | T | 0.0756 | 0.0036 | 1.54E-95 | LONP1 | 0.0865 | No | Yes |
| rs35586766 | 19_17392629_G_A | 19 | 17392629 | G | A | 0.0313 | 0.0036 | 4.48E-18 | ANKLE1 | 0.1103 | Yes | No |
| rs117176661 | 19_17445208_C_T | 19 | 17445208 | C | T | 0.0369 | 0.0048 | 1.32E-14 | ANO8; GTPBP3 | 0.0457 | No | Yes |
| rs2304128 | 19_19746151_G_T | 19 | 19746151 | G | T | 0.0226 | 0.0039 | 9.17E-09 | GMIP | 0.0805 | Yes | No |
| rs1065853 | 19_45413233_G_T | 19 | 45413233 | G | T | 0.0388 | 0.0039 | 1.59E-23 | APOE | 0.0626 | Yes | No |
| rs6105852 | 20_17955153_G_A | 20 | 17955153 | G | A | 0.0221 | 0.0021 | 1.41E-25 | MGME1 | 0.5179 | No | Yes |
| rs76599088 | 20_17968871_C_T | 20 | 17968871 | C | T | 0.0623 | 0.0082 | 3.16E-14 | MGME1 | 0.0139 | No | Yes |
| rs2245946 | 22_43528240_G_A | 22 | 43528240 | G | A | 0.0317 | 0.0023 | 9.38E-45 | MCAT | 0.663 | No | Yes |
| rs12148 | 22_50962208_T_G | 22 | 50962208 | T | G | -0.0139 | 0.0022 | 1.32E-10 | ODF3B | 0.6799 | No | Yes |

**S. Table 7.** Individual effects of MR variants on mtDNA-CN levels (UKBiobank) and 3-month mRS-based outcomes (GISCOME). Outcomes include ordinal mRS, poor functional outcome status (mRS 3-6 vs. 0-2), or mRS 2-6 vs. 0-1. Genetic effects for mtDNA-CN are expressed per 1 SD increase in mtDNA-CN levels. Genetic effects for mRS-based outcomes are expressed in terms of a log(odds) increase per additional alternative ("alt") allele. Accordingly, a positive beta indicates that the alternative allele increases risk of worse stroke outcome and vice versa.

| rsid | snpid | beta mtdna | se mtdna | pvalue mtdna | beta ordinal mRS | se ordinal mRS | pvalue ordinal mRS | beta mRS 3-6 vs 0-2 | se mRS 3-6 vs 0-2 | pvalue mRS 3-6 vs 0-2 | beta mRS 2-6 vs 0-1 | se mRS 2-6 vs 0-1 | pvalue mRS 2-6 vs 0-1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs3766744 | 1_12043717_G_A | -0.0179 | 0.0021 | 2.91E-17 | 0.0456 | 0.0343 | 0.1844 | 0.0546 | 0.0461 | 0.236 | 0.0595 | 0.0465 | 0.2009 |
| rs1342442 | 1_156466699_G_A | -0.0138 | 0.0022 | 5.56E-10 | -0.0604 | 0.0366 | 0.09821 | -0.0703 | 0.0487 | 0.1487 | -0.0861 | 0.05 | 0.08471 |
| rs11677402 | 2_74154975_G_T | 0.0251 | 0.0028 | 7.37E-20 | 0.0188 | 0.0473 | 0.691 | -0.0222 | 0.0577 | 0.7009 | -0.0155 | 0.0648 | 0.8115 |
| rs62641680 | 2_74166053_G_A | -0.0903 | 0.0063 | 4.63E-47 | 0.377 | 0.1212 | 0.001864 | 0.2712 | 0.1482 | 0.06714 | 0.3544 | 0.1743 | 0.04195 |
| rs74874677 | 2_74177777_A_G | -0.0819 | 0.0071 | 3.62E-31 | 0.2915 | 0.1247 | 0.01942 | 0.0832 | 0.1545 | 0.5901 | 0.2824 | 0.1765 | 0.1096 |
| rs6792510 | 3_48723302_G_C | -0.0126 | 0.0022 | 1.44E-08 | 0.0379 | 0.0361 | 0.2946 | 0.038 | 0.0443 | 0.3911 | 0.0686 | 0.0488 | 0.1599 |
| rs13088724 | 3_179152841_G_A | 0.0175 | 0.0026 | 7.70E-12 | 0.0447 | 0.0409 | 0.2741 | -0.0028 | 0.05 | 0.9556 | 0.1042 | 0.0557 | 0.06156 |
| rs5745582 | 6_33546498_C_T | 0.021 | 0.0028 | 3.46E-14 | -0.0659 | 0.042 | 0.1171 | -0.047 | 0.0566 | 0.4062 | -0.109 | 0.0568 | 0.05488 |
| rs2304693 | 7_45148667_G_A | 0.0181 | 0.0028 | 5.11E-11 | 0.067 | 0.0465 | 0.1493 | 0.1205 | 0.0563 | 0.03243 | 0.0454 | 0.0647 | 0.4825 |
| rs2322718 | 8_27257787_T_G | 0.0136 | 0.0021 | 1.53E-10 | 0.0005 | 0.0344 | 0.9877 | -0.0246 | 0.042 | 0.5578 | -0.0187 | 0.0469 | 0.6906 |
| rs12247015 | 10_60145079_A_G | 0.0337 | 0.0021 | 1.28E-55 | 0.0275 | 0.0348 | 0.4298 | 0.0182 | 0.0428 | 0.6697 | 0.0089 | 0.0471 | 0.8494 |
| rs57066921 | 10_102752345_T_G | -0.1081 | 0.0081 | 8.98E-41 | -0.0033 | 0.1135 | 0.9771 | 0.1235 | 0.1417 | 0.3834 | 0.1082 | 0.1548 | 0.4846 |
| rs701834 | 10_102761801_C_T | -0.0254 | 0.0026 | 1.12E-22 | -0.0061 | 0.0426 | 0.8862 | -0.043 | 0.0519 | 0.4069 | 0.0099 | 0.0582 | 0.8646 |
| rs11596235 | 10_104391034_C_T | 0.0174 | 0.0023 | 2.16E-14 | 0.0012 | 0.0376 | 0.9752 | -0.0032 | 0.0458 | 0.9436 | 0.041 | 0.0509 | 0.4196 |
| rs1127787 | 12_27867727_G_A | -0.0159 | 0.0028 | 1.49E-08 | 0.0386 | 0.0464 | 0.4061 | -0.008 | 0.0575 | 0.8898 | 0.1172 | 0.0642 | 0.06802 |
| rs12426673 | 12_109490296_G_T | -0.0141 | 0.0021 | 4.03E-11 | -0.019 | 0.0344 | 0.5803 | -0.0431 | 0.0421 | 0.3065 | -0.0313 | 0.0466 | 0.5018 |
| rs1760940 | 14_20938251_A_C | 0.0263 | 0.0024 | 5.20E-27 | -0.0214 | 0.0418 | 0.6082 | -0.0203 | 0.0515 | 0.6932 | -0.0234 | 0.0563 | 0.6776 |
| rs3889402 | 17_18167397_G_T | 0.026 | 0.0034 | 2.60E-14 | -0.0166 | 0.0596 | 0.7805 | -0.0636 | 0.0732 | 0.3847 | -0.01 | 0.0818 | 0.9031 |
| rs62072546 | 17_18253839_C_T | -0.0149 | 0.0025 | 2.95E-09 | 0.0068 | 0.0413 | 0.8695 | -0.009 | 0.0507 | 0.8591 | -0.008 | 0.057 | 0.8885 |
| rs17850455 | 17_62476451_C_G | 0.091 | 0.0104 | 1.81E-18 | -0.2642 | 0.1722 | 0.125 | -0.3578 | 0.2139 | 0.09436 | 0.1666 | 0.2532 | 0.5105 |
| rs11085147 | 19_5711930_C_T | 0.0756 | 0.0036 | 1.54E-95 | -0.1133 | 0.0607 | 0.06193 | -0.1782 | 0.0763 | 0.01952 | -0.1012 | 0.0848 | 0.2332 |
| rs117176661 | 19_17445208_C_T | 0.0369 | 0.0048 | 1.32E-14 | 0.0964 | 0.0896 | 0.2821 | -0.0449 | 0.1135 | 0.6922 | 0.0837 | 0.1207 | 0.4878 |
| rs6105852 | 20_17955153_G_A | 0.0221 | 0.0021 | 1.41E-25 | -0.0646 | 0.0346 | 0.06158 | -0.0843 | 0.0423 | 0.0466 | -0.061 | 0.047 | 0.1945 |
| rs76599088 | 20_17968871_C_T | 0.0623 | 0.0082 | 3.16E-14 | -0.1168 | 0.1697 | 0.4911 | 0.0687 | 0.2099 | 0.7436 | -0.3896 | 0.2448 | 0.1116 |
| rs2245946 | 22_43528240_G_A | 0.0317 | 0.0023 | 9.38E-45 | -0.0177 | 0.0369 | 0.6318 | -0.0022 | 0.0504 | 0.9656 | -0.0136 | 0.0495 | 0.7842 |
| rs12148 | 22_50962208_T_G | -0.0139 | 0.0022 | 1.32E-10 | -0.002 | 0.036 | 0.9554 | 0.021 | 0.0439 | 0.6328 | 0.0302 | 0.0485 | 0.5335 |

**S. Table 8.** Main MR results for the association between genetically determined mtDNA-CN levels and 3-month post-stroke outcomes. Results are expressed per 1 SD decrease in genetically predicted mtDNA-CN.

| Analysis Parameters | | | | | Main MR Results | | | | | | Sensitivity Analyses | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNPS included | $R^2$ explained | F-statistic | 3-month Stroke Outcome | MR Method | Beta | SE | OR | LCI | UCI | P | MR-PRESSO Global Test P | Q | Q df | Q P | MR-EGGER intercept | MR-EGGER intercept se | MR-EGGER intercept P |
| 1_12043717_g_a; 1_156466699_g_a; 10_102752345_t_g; 10_102761801_c_t; 10_104391034_c_t; 10_60145079_a_g; 12_109490296_g_t; 12_27867727_g_a; 14_20938251_a_c; 17_18167397_g_t; 17_18253839_c_t; 17_62476451_c_g; 19_17445208_c_t; 19_5711930_c_t; 2_74154975_g_t; 2_74166053_g_a; 2_74177777_a_g; 20_17955153_g_a; 20_17968871_c_t; 22_43528240_g_a; 22_50962208_t_g; 3_179152841_g_a; 3_48723302_g_c; 6_33546498_c_t; 7_45148667_g_a; 8_27257787_t_g | 0.0068 | 100 | Ordinal mRS | Inverse variance weighted | 0.85 | 0.37 | 2.35 | 1.13 | 4.90 | 0.02 | 0.12 | 32.96 | 25 | 0.13 | NA | NA | NA |
| | | | Ordinal mRS | Weighted median | 0.62 | 0.50 | 1.86 | 0.69 | 4.97 | 0.22 | 0.12 | NA | NA | NA | NA | NA | NA |
| | | | Ordinal mRS | MR Egger | 1.86 | 0.67 | 6.44 | 1.75 | 23.75 | 0.01 | 0.12 | 29.06 | 24 | 0.22 | -0.03 | 0.02 | 0.09 |
| | | | mRS 2-6 vs. 0-1 | Inverse variance weighted | 1.00 | 0.41 | 2.72 | 1.22 | 6.05 | 0.01 | 0.30 | 23.99 | 25 | 0.52 | NA | NA | NA |
| | | | mRS 2-6 vs. 0-1 | Weighted median | 1.11 | 0.63 | 3.04 | 0.89 | 10.40 | 0.08 | 0.30 | NA | NA | NA | NA | NA | NA |
| | | | mRS 2-6 vs. 0-1 | MR Egger | 2.14 | 0.75 | 8.48 | 1.93 | 37.24 | 0.01 | 0.30 | 20.77 | 24 | 0.65 | 0.04 | 0.02 | 0.09 |
| | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | Inverse variance weighted | 0.99 | 0.48 | 2.68 | 1.05 | 6.86 | 0.04 | 0.51 | 28.56 | 25 | 0.28 | NA | NA | NA |
| | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | Weighted median | 0.97 | 0.66 | 2.63 | 0.73 | 9.54 | 0.14 | 0.51 | NA | NA | NA | NA | NA | NA |
| | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | MR Egger | 1.71 | 0.90 | 5.55 | 0.95 | 32.49 | 0.07 | 0.51 | 27.51 | 24 | 0.28 | 0.02 | 0.02 | 0.35 |

**S. Table 9.** Sensitivity analyses showing MR results for the association between genetically determined blood cell traits and 3-month post-stroke outcomes. Results are expressed per 1 SD increase in genetically predicted blood cell counts.

| Analysis Parameters | | | | | | Main MR Results | | | | | | Sensitivity Analyses | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # SNPs | R2 explained | F-statistic | Exposure | 3-month Stroke Outcome | MR Method | Beta | SE | OR | LCI | UCI | P | MR-PRESSO Global Test P | Q | Q df | Q P | MR-EGGER intercept | MR-EGGER intercept se | MR-EGGER intercept P |
| 694 | 0.11 | 96.99 | Neutrophil Count | Ordinal mRS | Inverse variance weighted | 0.00 | 0.07 | 1.00 | 0.87 | 1.15 | 0.99 | 0.79 | 689.17 | 693 | 0.53 | NA | NA | NA |
| | | | | Ordinal mRS | Weighted median | 0.05 | 0.12 | 1.05 | 0.83 | 1.34 | 0.67 | 0.79 | NA | NA | NA | NA | NA | NA |
| | | | | Ordinal mRS | MR Egger | 0.14 | 0.15 | 1.16 | 0.86 | 1.56 | 0.34 | 0.79 | 688.03 | 692 | 0.54 | 0.00 | 0.00 | 0.29 |
| 696 | 0.11 | 99.74 | | mRS 2-6 vs. 0-1 | Inverse variance weighted | 0.01 | 0.10 | 1.01 | 0.83 | 1.22 | 0.95 | 0.72 | 676.18 | 695 | 0.69 | NA | NA | NA |
| | | | | mRS 2-6 vs. 0-1 | Weighted median | -0.14 | 0.17 | 0.87 | 0.63 | 1.21 | 0.40 | 0.72 | NA | NA | NA | NA | NA | NA |
| | | | | mRS 2-6 vs. 0-1 | MR Egger | 0.02 | 0.21 | 1.02 | 0.68 | 1.53 | 0.93 | 0.72 | 676.18 | 694 | 0.68 | 0.00 | 0.00 | 0.95 |
| 694 | 0.11 | 100.47 | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | Inverse variance weighted | 0.08 | 0.09 | 1.08 | 0.90 | 1.29 | 0.40 | 0.77 | 664.76 | 693 | 0.77 | NA | NA | NA |
| | | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | Weighted median | 0.03 | 0.15 | 1.03 | 0.77 | 1.39 | 0.83 | 0.77 | NA | NA | NA | NA | NA | NA |
| | | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | MR Egger | 0.13 | 0.19 | 1.14 | 0.79 | 1.65 | 0.48 | 0.77 | 664.64 | 692 | 0.77 | 0.00 | 0.00 | 0.73 |
| 889 | 0.14 | 99.67 | White Blood Cell Count | Ordinal mRS | Inverse variance weighted | 0.07 | 0.07 | 1.07 | 0.94 | 1.22 | 0.31 | 0.57 | 881.71 | 888 | 0.55 | NA | NA | NA |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Ordinal mRS | Weighted median | 0.10 | 0.11 | 1.11 | 0.89 | 1.38 | 0.35 | 0.57 | NA | NA | NA | NA | NA | NA |
| | | | | Ordinal mRS | MR Egger | 0.19 | 0.14 | 1.21 | 0.92 | 1.59 | 0.17 | 0.57 | 880.67 | 887 | 0.55 | 0.00 | 0.00 | 0.31 |
| 895 | 0.14 | 106.15 | | mRS 2-6 vs. 0-1 | Inverse variance weighted | 0.06 | 0.09 | 1.06 | 0.89 | 1.26 | 0.52 | 0.99 | 811.02 | 894 | 0.98 | NA | NA | NA |
| | | | | mRS 2-6 vs. 0-1 | Weighted median | -0.01 | 0.15 | 0.99 | 0.73 | 1.33 | 0.94 | 0.99 | NA | NA | NA | NA | NA | NA |
| | | | | mRS 2-6 vs. 0-1 | MR Egger | 0.01 | 0.19 | 1.01 | 0.70 | 1.46 | 0.96 | 0.99 | 810.94 | 893 | 0.98 | 0.00 | 0.00 | 0.78 |
| 891 | 0.14 | 106.20 | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | Inverse variance weighted | 0.07 | 0.08 | 1.08 | 0.92 | 1.26 | 0.37 | 0.73 | 864.28 | 890 | 0.73 | NA | NA | NA |
| | | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | Weighted median | 0.08 | 0.14 | 1.09 | 0.82 | 1.44 | 0.56 | 0.73 | NA | NA | NA | NA | NA | NA |
| | | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | MR Egger | 0.19 | 0.17 | 1.21 | 0.86 | 1.70 | 0.26 | 0.73 | 863.66 | 889 | 0.72 | 0.00 | 0.00 | 0.43 |
| 1073 | 0.23 | 154.61 | Platelet Count | Ordinal mRS | Inverse variance weighted | 0.02 | 0.05 | 1.02 | 0.92 | 1.13 | 0.72 | 0.14 | 1125.03 | 1072 | 0.13 | NA | NA | NA |
| | | | | Ordinal mRS | Weighted median | 0.05 | 0.09 | 1.05 | 0.87 | 1.25 | 0.62 | 0.14 | NA | NA | NA | NA | NA | NA |
| | | | | Ordinal mRS | MR Egger | 0.05 | 0.10 | 1.06 | 0.87 | 1.27 | 0.57 | 0.14 | 1124.83 | 1071 | 0.12 | 0.00 | 0.00 | 0.66 |
| 1074 | 0.23 | 154.60 | | mRS 2-6 vs. 0-1 | Inverse variance weighted | 0.04 | 0.07 | 1.04 | 0.90 | 1.19 | 0.61 | 0.60 | 1059.93 | 1073 | 0.61 | NA | NA | NA |
| | | | | mRS 2-6 vs. 0-1 | Weighted median | -0.03 | 0.12 | 0.97 | 0.77 | 1.22 | 0.78 | 0.60 | NA | NA | NA | NA | NA | NA |
| | | | | mRS 2-6 vs. 0-1 | MR Egger | 0.02 | 0.13 | 1.02 | 0.79 | 1.32 | 0.86 | 0.60 | 1059.92 | 1072 | 0.60 | 0.00 | 0.00 | 0.91 |
| 1071 | 0.23 | 154.69 | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | Inverse variance weighted | 0.02 | 0.06 | 1.02 | 0.90 | 1.16 | 0.74 | 0.64 | 1053.28 | 1070 | 0.64 | NA | NA | NA |
| | | | | Poor Functional Outcome | Weighted median | 0.14 | 0.11 | 1.15 | 0.93 | 1.42 | 0.21 | 0.64 | NA | NA | NA | NA | NA | NA |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | (mRS 3-6 vs. 0-2) | | | | | | | | | | | | | | |
| | | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | MR Egger | 0.08 | 0.12 | 1.09 | 0.86 | 1.37 | 0.47 | 0.64 | 1052.87 | 1069 | 0.63 | 0.00 | 0.00 | 0.52 |
| 874 | 0.14 | 108.33 | Lymphocyte Count | Ordinal mRS | Inverse variance weighted | -0.01 | 0.06 | 0.99 | 0.87 | 1.12 | 0.84 | 0.35 | 890.72 | 873 | 0.33 | NA | NA | NA |
| | | | | Ordinal mRS | Weighted median | 0.11 | 0.12 | 1.12 | 0.89 | 1.40 | 0.34 | 0.35 | NA | NA | NA | NA | NA | NA |
| | | | | Ordinal mRS | MR Egger | 0.23 | 0.13 | 1.25 | 0.97 | 1.62 | 0.08 | 0.35 | 886.23 | 872 | 0.36 | 0.01 | 0.00 | 0.04 |
| 877 | 0.14 | 108.13 | | mRS 2-6 vs. 0-1 | Inverse variance weighted | -0.08 | 0.09 | 0.93 | 0.78 | 1.10 | 0.38 | 0.60 | 867.88 | 876 | 0.57 | NA | NA | NA |
| | | | | mRS 2-6 vs. 0-1 | Weighted median | 0.07 | 0.15 | 1.08 | 0.81 | 1.43 | 0.61 | 0.60 | NA | NA | NA | NA | NA | NA |
| | | | | mRS 2-6 vs. 0-1 | MR Egger | 0.07 | 0.18 | 1.08 | 0.76 | 1.52 | 0.68 | 0.60 | 866.95 | 875 | 0.57 | 0.00 | 0.00 | 0.34 |
| 870 | 0.14 | 107.57 | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | Inverse variance weighted | -0.09 | 0.08 | 0.91 | 0.78 | 1.07 | 0.27 | 0.96 | 800.53 | 869 | 0.95 | NA | NA | NA |
| | | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | Weighted median | -0.04 | 0.13 | 0.96 | 0.74 | 1.25 | 0.78 | 0.96 | NA | NA | NA | NA | NA | NA |
| | | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | MR Egger | 0.16 | 0.17 | 1.17 | 0.85 | 1.62 | 0.33 | 0.96 | 797.54 | 868 | 0.96 | -0.01 | 0.00 | 0.08 |
| 195 | 0.03 | 61.28 | Neutrophil to Lymphocyte Ratio | Ordinal mRS | Inverse variance weighted | 0.02 | 0.15 | 1.02 | 0.77 | 1.36 | 0.87 | 0.05 | 228.12 | 194 | 0.05 | NA | NA | NA |
| | | | | Ordinal mRS | Weighted median | 0.13 | 0.22 | 1.13 | 0.74 | 1.73 | 0.56 | 0.05 | NA | NA | NA | NA | NA | NA |
| | | | | Ordinal mRS | MR Egger | -0.15 | 0.40 | 0.86 | 0.40 | 1.87 | 0.71 | 0.05 | 227.87 | 193 | 0.04 | 0.00 | 0.01 | 0.64 |
| 197 | 0.03 | 61.33 | | mRS 2-6 vs. 0-1 | Inverse variance weighted | -0.11 | 0.18 | 0.90 | 0.63 | 1.28 | 0.54 | 0.62 | 190.56 | 196 | 0.60 | NA | NA | NA |
| | | | | mRS 2-6 vs. 0-1 | Weighted median | -0.16 | 0.29 | 0.86 | 0.48 | 1.52 | 0.59 | 0.62 | NA | NA | NA | NA | NA | NA |
| | | | | mRS 2-6 vs. 0-1 | MR Egger | -0.06 | 0.47 | 0.94 | 0.37 | 2.39 | 0.90 | 0.62 | 190.54 | 195 | 0.58 | 0.00 | 0.01 | 0.91 |
| 196 | 0.03 | 61.60 | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | Inverse variance weighted | 0.14 | 0.17 | 1.15 | 0.83 | 1.59 | 0.41 | 0.47 | 197.10 | 195 | 0.44 | NA | NA | NA |
| | | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | Weighted median | 0.09 | 0.26 | 1.10 | 0.65 | 1.84 | 0.73 | 0.47 | NA | NA | NA | NA | NA | NA |
| | | | | Poor Functional Outcome (mRS 3-6 vs. 0-2) | MR Egger | -0.15 | 0.45 | 0.86 | 0.35 | 2.10 | 0.75 | 0.47 | 196.64 | 194 | 0.43 | 0.01 | 0.01 | 0.50 |

**S. Table 10.** STREGA checklist.

| Section Number | Section(s) | Item | Place Addressed |
|---|---|---|---|
| 1 | TITLE and ABSTRACT | a) Indicate the study's design with a commonly used term in the title or the abstract | Title Page and Abstract |
| | | b) Provide in the abstract an informative and balanced summary of what was done and what was found | Abstract |
| 2 | BACKGROUND and RATIONALE | Explain the scientific background and rationale for the investigation being reported. | Introduction |
| 3 | OBJECTIVES | State specific objectives, including any prespecified hypotheses. State if the study is the first report of a genetic association, a replication effort, or both. | Introduction |
| 4 | STUDY DESIGN | Present key elements of study design early in the paper | Introduction |
| 5 | SETTING | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | Methods ("INTERSTROKE") |

| | | | |
|---|---|---|---|
| 6 | ELIGIBILITY CRITERIA | a) Cohort study – Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. Case-control study – Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. Cross-sectional study – Give the eligibility criteria, and the sources and methods of selection of participants. Give information on the criteria and methods for selection of subsets of participants from a larger study, when relevant. | Methods ("INTERSTROKE"), Results ("Baseline Characteristics of INTERSTROKE cases"), Supplementary Figure 1 |
| | | b) Cohort study – For matched studies, give matching criteria and number of exposed and unexposed. Case-control study – For matched studies, give matching criteria and the number of controls per case. | NA |
| 7 | VARIABLES | a) Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | Methods ("MtDNA-CN Measurement and Quality Control", "Statistical Analysis","Mendelian Randomization") |
| | | b) Clearly define genetic exposures (genetic variants) using a widely-used nomenclature system. Identify variables likely to be associated with population stratification (confounding by ethnic origin). | Supplementary Tables 6-8 |
| 8 | DATA SOURCES / MEASUREMENT | a) For each variable of interest give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group. Give information separately for for exposed and unexposed groups if applicable. | Methods ("MtDNA-CN Measurement and Quality Control", "Statistical Analysis","Mendelian Randomization") |
| | | b) Describe laboratory methods, including source and storage of DNA, genotyping methods and platforms (including the allele calling algorithm used, and its version), error rates and call rates. State the laboratory / centre where genotyping was done. Describe comparability of laboratory methods if there is more than one group. Specify whether genotypes were assigned using all of the data from the study simultaneously or in smaller batches. | Methods ("MtDNA-CN Measurement and Quality Control") |
| 9 | BIAS | Describe any efforts to address potential sources of bias | Methods ("Statistical Analysis") |
| 10 | STUDY SIZE | Explain how the study size was arrived at | Supplementary Figure 1 |
| 11 | QUANTITATIVE VARIABLES | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why. If applicable, describe how effects of treatment were dealt with. | Methods ("MtDNA-CN Measurement and Quality Control") |
| 12 | STATISTICAL METHODS | Describe any methods used to assess the assumptions or justify their validity. | Methods ("MtDNA-CN Measurement and Quality Control") |
| | | a) Describe all statistical methods, including those used to control for confounding. State software version used and options (or settings) chosen. | Methods ("Statistical Analysis") |
| | | b) Describe any methods used to examine subgroups and interactions | Methods ("Statistical Analysis") |
| | | c) Explain how missing data were addressed | Methods ("Statistical Analysis") |
| | | d) If applicable, explain how loss to follow-up was addressed | NA |
| | | e) Describe any sensitivity analyses | Methods ("Statistical Analysis") |
| | | f) State whether Hardy-Weinberg equilibrium was considered and, if so, how. | NA |
| | | g) Describe any methods used for inferring genotypes or haplotypes | NA |
| | | h) Describe any methods used to assess or address population stratification. | NA |

| | | | |
|---|---|---|---|
| | | i) Describe any methods used to address multiple comparisons or to control risk of false positive findings. | Methods ("Statistical Analysis"; Page 8) |
| | | j) Describe any methods used to address and correct for relatedness among subjects | NA |
| 13 | PARTICIPANTS | a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed. Give information separately for for exposed and unexposed groups if applicable. Report numbers of individuals in whom genotyping was attempted and numbers of individuals in whom genotyping was successful. | Supplementary Figure 1 |
| | | b) Give reasons for non-participation at each stage | Supplementary Figure 1 |
| | | c) Consider use of a flow diagram | Supplementary Figure 1 |
| 14 | DESCRIPTIVE DATA | a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. Give information separately for exposed and unexposed groups if applicable. Consider giving information by genotype | Results Table 1 |
| | | b) Indicate number of participants with missing data for each variable of interest | Results Table 1 |
| | | c) Cohort study – Summarize follow-up time, e.g. average and total amount. | Results |
| 15 | OUTCOME DATA | Cohort study Report numbers of outcome events or summary measures over time.Give information separately for exposed and unexposed groups if applicable. Report outcomes (phenotypes) for each genotype category over time Case-control study – Report numbers in each exposure category, or summary measures of exposure.Give information separately for cases and controls . Report numbers in each genotype category. Cross-sectional study – Report numbers of outcome events or summary measures. Give information separately for exposed and unexposed groups if applicable. Report outcomes (phenotypes) for each genotype category | Results ("Lower mtDNA-CN is associated with poor stroke prognosis at 1-month") |
| 16 | MAIN RESULTS | a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | Methods ("Statistical Analysis") |
| | | b) Report category boundaries when continuous variables were categorized | Methods ("Statistical Analysis") |
| | | c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | NA |
| | | d) Report results of any adjustments for multiple comparisons | NA |
| 17 | OTHER ANALYSES | a) Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses | Results ("Lower mtDNA-CN is associated with poor stroke prognosis at 1-month"), Figure 3 |
| | | b) If numerous genetic exposures (genetic variants) were examined, summarize results from all analyses undertaken. | NA |
| | | c) If detailed results are available elsewhere, state how they can be accessed. | Supplementary Tables 1 to 5 |
| 18 | KEY RESULTS | Summarise key results with reference to study objectives | Discussion |
| 19 | LIMITATIONS | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias. | Discussion |
| 20 | INTERPRETATION | Give a cautious overall interpretation considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence. | Discussion + Conclusion |
| 21 | GENERALISABILITY | Discuss the generalisability (external validity) of the study results | Discussion + Conclusion |
| 22 | FUNDING | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | Funding Sources |

**S. Table 11.** STROBE-MR checklist.

| Section Number | Section(s) | Item | Place Addressed |
|---|---|---|---|
| 1 | TITLE and ABSTRACT | Indicate Mendelian randomization as the study's design in the title and/or abstract | Title page and abstract |
| 2 | INTRODUCTION: Background | Explain the scientific background and rationale for the reported study. Is causality between exposure and outcome plausible? Justify why MR is a helpful method to address the study question. | Introduction |
| 3 | INTRODUCTION: Objectives | State specific objectives clearly, including pre-specified causal hypotheses (if any). | Introduction |
| 4 | METHODS: Study Design + Data Sources | Present key elements of study design early in the paper. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following: | Methods ("Mendelian Randomization") |
| | | a) Describe the study design and the underlying population from which it was drawn. Describe also the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, if available. | Methods ("Mendelian Randomization") |
| | | b) Give the eligibility criteria, and the sources and methods of selection of participants. | Methods ("Mendelian Randomization") |
| | | c) Explain how the analyzed sample size was arrived at. | Methods ("Mendelian Randomization") |
| | | d) Describe measurement, quality and selection of genetic variants. | Methods ("Mendelian Randomization") |
| | | e) For each exposure, outcome and other relevant variables, describe methods of assessment and, in the case of diseases, the diagnostic criteria used. | Methods ("Mendelian Randomization") |
| | | f) Provide details of ethics committee approval and participant informed consent, if relevant. | NA |
| 5 | METHODS: Assumptions | Explicitly state assumptions for the main analysis (e.g. relevance, exclusion, independence, homogeneity) as well assumptions for any additional or sensitivity analysis. | Methods ("Mendelian Randomization"); Supplementary Methods ("Mendelian Randomization Instrument Selection") |
| 6 | METHODS: Statistical Methods: main analysis | a) Describe how quantitative variables were handled in the analyses (i.e., scale, units, model). | Methods ("Mendelian Randomization") |
| | | b) Describe the process for identifying genetic variants and weights to be included in the analyses (i.e, independence and model). Consider a flow diagram. | Methods ("Mendelian Randomization"); Supplementary Methods ("Mendelian Randomization Instrument Selection") |
| | | c) Describe the MR estimator, e.g. two-stage least squares, Wald ratio, and related statistics. Detail the included covariates and, in case of two-sample MR, whether the same covariate set was used for adjustment in the two samples. | Methods ("Mendelian Randomization") |
| | | d) Explain how missing data were addressed. | Methods ("Mendelian Randomization"); Supplementary Methods ("Mendelian Randomization Instrument Selection") |
| | | e) If applicable, say how multiple testing was dealt with. | NA |
| 7 | METHODS: Assesment of Assumptions | Describe any methods used to assess the assumptions or justify their validity. | Methods ("Mendelian Randomization") |
| 8 | METHODS: Sensitivity Analyses | Describe any sensitivity analyses or additional analyses performed. | Methods ("Mendelian Randomization") |

| | | | |
|---|---|---|---|
| 9 | METHODS: Software & Pre-registration | a) Name statistical software and package(s), including version and settings used. | Methods ("Mendelian Randomization") |
| | | b) State whether the study protocol and details were pre-registered (as well as when and where). | NA |
| 10 | RESULTS: Descriptive Data | a) Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow-diagram. | Methods ("Mendelian Randomization") |
| | | b) Report summary statistics for phenotypic exposure(s), outcome(s) and other relevant variables (e.g. means, standard deviations, proportions). | Methods ("Mendelian Randomization") |
| | | c) If the data sources include meta-analyses of previous studies, provide the number of studies, their reported ancestry, if available, and assessments of heterogeneity across these studies. Consider using a supplementary table for each data source. | Methods ("Mendelian Randomization") |
| | | d) For two-sample Mendelian randomization:<br>i. Provide information on the similarity of the genetic variant-exposure associations between the exposure and outcome samples. | NA |
| | | ii. Provide information on extent of sample overlap between the exposure and outcome data sources. | Methods ("Mendelian Randomization") |
| 11 | RESULTS: Main Results | a) Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale (e.g. comparing 25th and 75th percentile of allele count or genetic risk score, if individual-level data available). | Supplementary Table 7 |
| | | b) Report causal effect estimate between exposure and outcome, and the measures of uncertainty from the MR analysis. Use an intuitive scale, such as odds ratio, or relative risk, per standard deviation difference. | Results ("Low mtDNA-CN is a putative causal risk factor for 3-month stroke outcomes"), Figure 4 |
| | | c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time-period. | NA |
| | | d) Consider any plots to visualize results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure). | Results Figure 4 |
| 12 | RESULTS: Assessment of Assumptions | a) Assess the validity of the assumptions. | Results ("Low mtDNA-CN is a putative causal risk factor for 3-month stroke outcomes"), Supplementary Tables 8 and 9 |
| | | b) Report any additional statistics (e.g., assessments of heterogeneity, such as I2 , Q statistic). | Results ("Low mtDNA-CN is a putative causal risk factor for 3-month stroke outcomes"), Supplementary Tables 8 and 9 |
| 13 | RESULTS: Sensitivity + Additional Analyses | a) Use sensitivity analyses to assess the robustness of the main results to violations of the assumptions. | Results ("Low mtDNA-CN is a putative causal risk factor for 3-month stroke outcomes"), Supplementary Tables 8 and 9 |
| | | b) Report results from other sensitivity analyses (e.g., replication study with different dataset, analyses of subgroups, validation of instrument(s), simulations, etc.). | Results ("Low mtDNA-CN is a putative causal risk factor for 3-month stroke outcomes"), Figure 4, Supplementary Table 9 |
| | | c) Report any assessment of direction of causality (e.g., bidirectional MR). | NA |
| | | d) When relevant, report and compare with estimates from non-MR analyses. | Discussion |

| | | | |
|---|---|---|---|
| | | e) Consider any additional plots to visualize results (e.g., leave-one-out analyses). | NA |
| 14 | DISCUSSION: Key Results | Summarize key results with reference to study objectives. | Discussion |
| 15 | DISCUSSION: Limitations | Discuss limitations of the study, taking into account the validity of the MR assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias, and any efforts to address them. | Discussion |
| 16 | DISCUSSION: Interpretation | a) Give a cautious overall interpretation of results considering objectives and limitations. Compare with results from other relevant studies. | Discussion + Conclusion |
| | | b) Discuss underlying biological mechanisms that could be modelled by using the genetic variants to assess the relationship between the exposure and the outcome. | Discussion |
| | | c) Discuss whether the results have clinical or policy relevance, and whether interventions could have the same size effect. | Discussion + Conclusion |
| 17 | DISCUSSION: Generalizability | Discuss the generalizability of the study results (a) to other populations (i.e. external validity), (b) across other exposure periods/timings, and (c) across other levels of exposure. | Discussion |
| 18 | OTHER INFO: Funding | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study or studies on which the present article is based. | Funding Sources |
| 19 | OTHER INFO: Data + Data Sharing | Present data used to perform all analyses or report where and how the data can be accessed. State whether statistical code is publicly accessible and if so, where. | Data Availability |
| 20 | OTHER INFO: Conflicts of Interest | All authors should declare all potential conflicts of interest. | Conflicts of Interest |

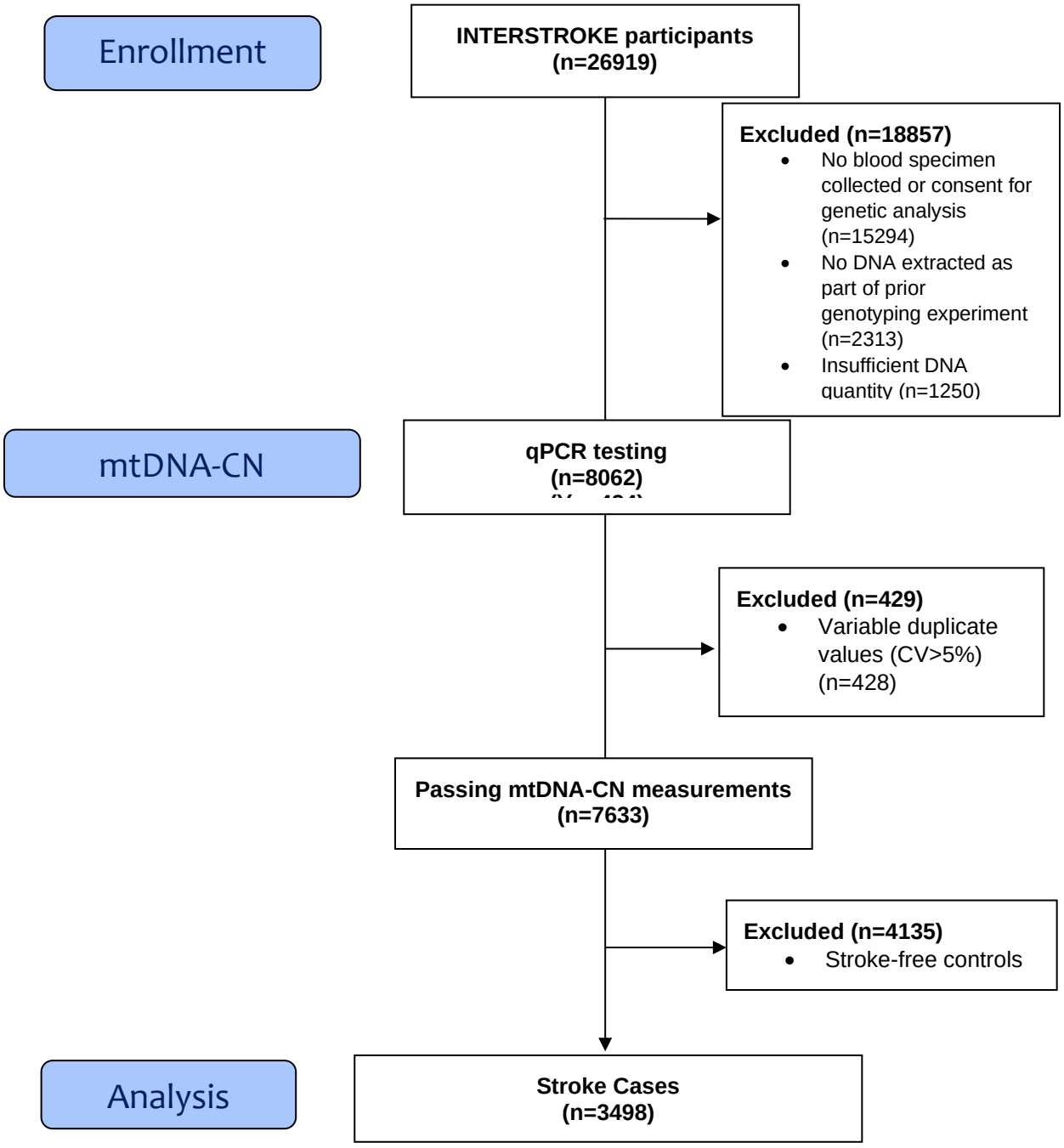
## Supplementary References


1. Fazzini F, Schöpf B, Blatzer M, et al. Plasmid-normalized quantification of relative mitochondrial DNA copy number. 2018;(May):1-11. doi:10.1038/s41598-018-33684-5
2. Chong M, Mohammadi-Shemirani P, Perrot N, et al. GWAS and ExWAS of blood Mitochondrial DNA copy number identifies 73 loci and highlights a potential causal role in dementia. *medRxiv*. Published online 2021. doi:10.1101/2021.04.08.21255031
3. Rath S, Sharma R, Gupta R, et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res*. 2021;49(D1):D1541-D1547. doi:10.1093/nar/gkaa1011
4. Kamat MA, Blackshaw JA, Young R, et al. PhenoScanner V2: an expanded tool for

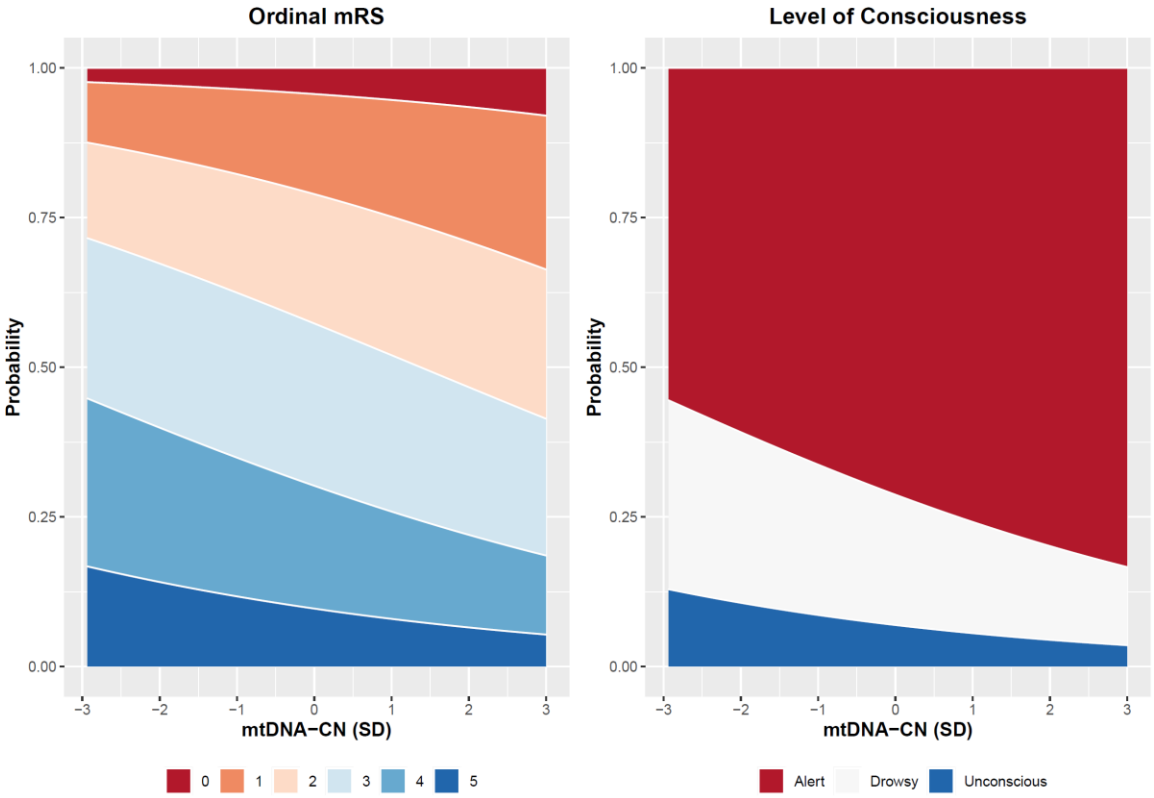
searching human genotype-phenotype associations. *Bioinformatics*. 2019;35(22):4851-4853. doi:10.1093/bioinformatics/btz469

5. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole genome regression for quantitative and binary traits.

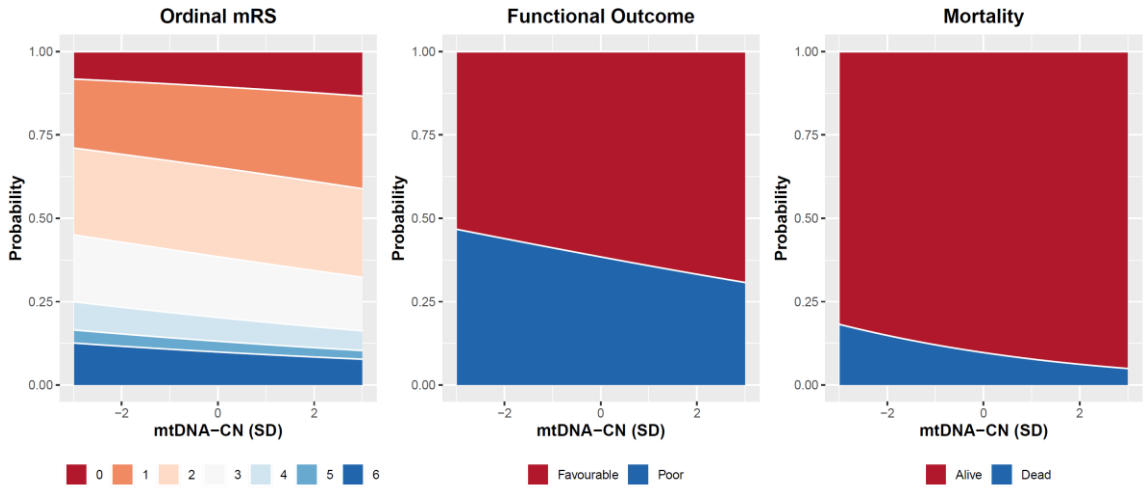
## Supplementary Figures



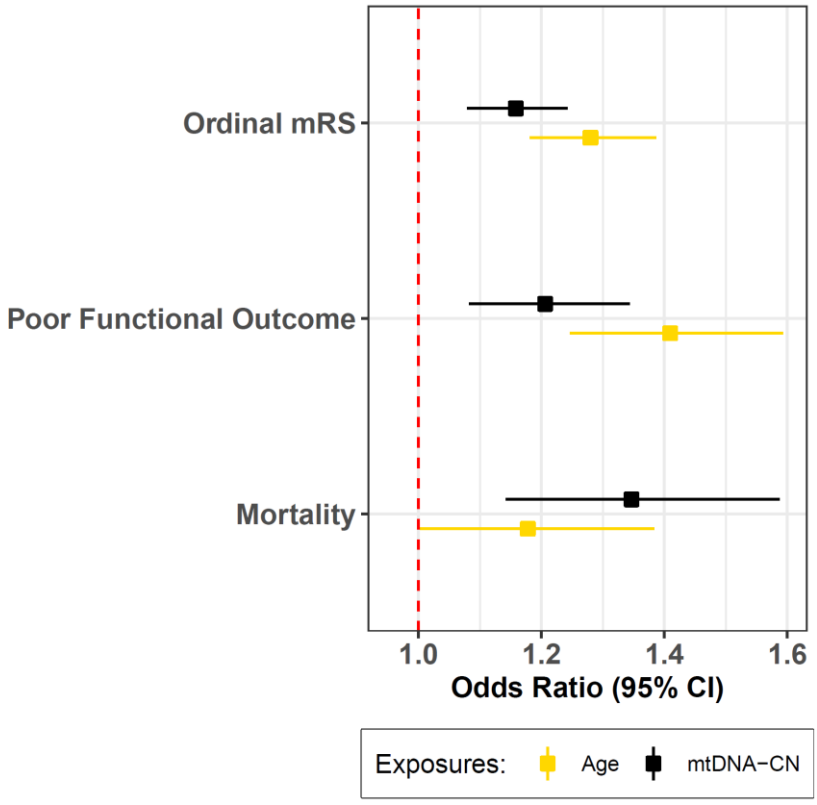

**Supplementary Figure 1.** Participant flow chart.




**S. Figure 2.** Area charts illustrate the predicted model probabilities for baseline stroke severity strata as a function of mtDNA-CN (continuous variable).



**S. Figure 3.** Area charts illustrate the predicted model probabilities for 1-month post-stroke outcomes as a function of mtDNA-CN (continuous variable).




**S. Figure 4**. Comparison of effects for age and mtDNA-CN on 1-month stroke outcomes.