2019

Associations of mitochondrial and nuclear mitochondrial variants and genes with seven metabolic traits

Aldi T. Kraja

Charles Gu

Mary F. Feitosa

Shiow J. Lin

Lihua Wang

See next page for additional authors

Authors

Aldi T. Kraja, Charles Gu, Mary F. Feitosa, Shiow J. Lin, Lihua Wang, Mary J. Wojczynski, Ruteja A. Barve, Ping An, Thomas J. Baranski, E. Warwick Daw, Lisa de Las Fuentes, Latisha D. Love-Gregory, Christine Williams, Wei Yang, D. C. Rao, Victor G. Davila-Roman, Michael Province, and et al

ARTICLE

Associations of Mitochondrial and Nuclear Mitochondrial Variants and Genes with Seven Metabolic Traits

Aldi T. Kraja,[1,101,*] Chunyu Liu,[2,101] Jessica L. Fetterman,[3,101] Mariaelisa Graff,[4,101] Christian Theil Have,[5,101] Charles Gu,[6,101] Lisa R. Yanek,[7] Mary F. Feitosa,[1] Dan E. Arking,[8] Daniel I. Chasman,[9,10] Kristin Young,[4] Symen Ligthart,[11] W. David Hill,[12] Stefan Weiss,[13] Jian'an Luan,[14] Franco Giulianini,[9] Ruifang Li-Gao,[15] Fernando P. Hartwig,[16,17] Shiow J. Lin,[1] Lihua Wang,[1] Tom G. Richardson,[17] Jie Yao,[18] Eliana P. Fernandez,[11] Mohsen Ghanbari,[11] Mary K. Wojczynski,[1] Wen-Jane Lee,[19,20] Maria Argos,[21] Sebastian M. Armasu,[22] Ruteja A. Barve,[23] Kathleen A. Ryan,[24] Ping An,[1] Thomas J. Baranski,[25] Suzette J. Bielinski,[26] Donald W. Bowden,[27] Ulrich Broeckel,[28] Kaare Christensen,[29] Audrey Y. Chu,[9] Janie Corley,[12] Simon R. Cox,[12] Andre G. Uitterlinden,[30] Fernando Rivadeneira,[30] Cheryl D. Cropp,[31] E. Warwick Daw,[1] Diana van Heemst,[32] Lisa de las Fuentes,[33] He Gao,[34] Ioanna Tzoulaki,[34,35] Tarunveer S. Ahluwalia,[36] Renée de Mutsert,[15] Leslie S. Emery,[37] A. Mesut Erzurumluoglu,[38] James A. Perry,[24] Mao Fu,[24] Nita G. Forouhi,[14] Zhenglong Gu,[39] Yang Hai,[18] Sarah E. Harris,[40] Gibran Hemani,[17] Steven C. Hunt,[41,42] Marguerite R. Irvin,[43] Anna E. Jonsson,[5] Anne E. Justice,[4,44] Nicola D. Kerrison,[14] Nicholas B. Larson,[26] Keng-Hung Lin,[45] Latisha D. Love-Gregory,[46] Rasika A. Mathias,[7,47] Joseph H. Lee,[48] Matthias Nauck,[49] Raymond Noordam,[32] Ken K. Ong,[14]

(Author list continued on next page)

Mitochondria (MT), the major site of cellular energy production, are under dual genetic control by 37 mitochondrial DNA (mtDNA) genes and numerous nuclear genes (MT-nDNA). In the CHARGEmtDNA+ Consortium, we studied genetic associations of mtDNA and MT-nDNA associations with body mass index (BMI), waist-hip-ratio (WHR), glucose, insulin, HOMA-B, HOMA-IR, and HbA1c. This 45-cohort collaboration comprised 70,775 (insulin) to 170,202 (BMI) pan-ancestry individuals. Validation and imputation of mtDNA variants was followed by single-variant and gene-based association testing. We report two significant common variants, one in *MT-ATP6* associated ($p \leq 5E{-}04$) with WHR and one in the *D-loop* with glucose. Five rare variants in *MT-ATP6*, *MT-ND5*, and *MT-ND6* associated with BMI, WHR, or insulin. Gene-based meta-analysis identified *MT-ND3* associated with BMI ($p \leq 1E{-}03$). We considered 2,282 MT-nDNA candidate gene associations compiled from online summary results for our traits (20 unique studies with 31 dataset consortia's genome-wide associations [GWASs]). Of these, 109 genes associated ($p \leq 1E{-}06$) with at least 1 of our 7 traits. We assessed regulatory features of variants in the 109 genes, *cis*- and *trans*-gene expression regulation, and performed enrichment and protein-protein interactions analyses. Of the identified mtDNA and MT-nDNA genes, 79 associated with adipose measures, 49 with glucose/insulin, 13 with risk for type 2 diabetes, and 18 with cardiovascular disease, indicating for pleiotropic effects with health implications. Additionally, 21 genes related to cholesterol, suggesting additional important roles for the genes identified. Our results suggest that mtDNA and MT-nDNA genes and variants reported make important contributions to glucose and insulin metabolism, adipocyte regulation, diabetes, and cardiovascular disease.

Introduction

Mitochondria (MT) are double membraned organelles that generate the majority of cellular adenosine triphosphate (ATP) and metabolites and play a central role in human health.[1] MT reside within cells and contain a separate maternally inherited[2–5] ~16.6 kb circular mtDNA genome, consisting of 37 genes (Figure 1). The mtDNA encodes 13 core catalytic peptides that form the oxidative phosphorylation complexes (I, III-V) as well as the machinery needed

[1]Division of Statistical Genomics, Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, MO 63110, USA; [2]Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA; [3]Evans Department of Medicine and Whitaker Cardiovascular Institute, Boston University School of Medicine, Boston, MA 02118, USA; [4]Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27516, USA; [5]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2200, Denmark; [6]Division of Biostatistics, Washington University School of Medicine, St Louis, MO 63110, USA; [7]GeneSTAR Research Program, Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; [8]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [9]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA; [10]Harvard Medical School, Boston, MA 02115, USA; [11]Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3015 CE, the Netherlands; [12]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK; [13]Interfaculty Institute for Genetics and Functional Genomics, University Medicine and University of Greifswald, Greifswald 17475, Germany; [14]MRC Epidemiology Unit, University of Cambridge School of Clinical

(Affiliations continued on next page)

James Pankow,[50] Amit Patki,[51] Alison Pattie,[12] Astrid Petersmann,[49] Qibin Qi,[52] Rasmus Ribel-Madsen,[5,53,54] Rebecca Rohde,[4] Kevin Sandow,[18] Theresia M. Schnurr,[5] Tamar Sofer,[37,55] John M. Starr,[12,56] Adele M. Taylor,[12] Alexander Teumer,[57] Nicholas J. Timpson,[17] Hugoline G. de Haan,[15] Yujie Wang,[4] Peter E. Weeke,[58] Christine Williams,[1] Hongsheng Wu,[59] Wei Yang,[60] Donglin Zeng,[61] Daniel R. Witte,[62] Bruce S. Weir,[37] Nicholas J. Wareham,[14] Henrik Vestergaard,[5,36] Stephen T. Turner,[63] Christian Torp-Pedersen,[64] Evie Stergiakouli,[17] Wayne Huey-Herng Sheu,[65,66,67,68] Frits R. Rosendaal,[15] M. Arfan Ikram,[11] Oscar H. Franco,[11,69] Paul M. Ridker,[9,10] Thomas T. Perls,[70] Oluf Pedersen,[5] Ellen A. Nohr,[71] Anne B. Newman,[72] Allan Linneberg,[73,74,75] Claudia Langenberg,[14] Tuomas O. Kilpeläinen,[5] Sharon L.R. Kardia,[76] Marit E. Jørgensen,[36] Torben Jørgensen,[77,78,79] Thorkild I.A. Sørensen,[80] Georg Homuth,[13] Torben Hansen,[5] Mark O. Goodarzi,[81] Ian J. Deary,[12] Cramer Christensen,[82] Yii-Der Ida Chen,[18] Aravinda Chakravarti,[83] Ivan Brandslund,[84,85] Klaus Bonnelykke,[86,87] Kent D. Taylor,[18] James G. Wilson,[88] Santiago Rodriguez,[17] Gail Davies,[12] Bernardo L. Horta,[16] Bharat Thyagarajan,[89] D.C. Rao,[6] Niels Grarup,[5] Victor G. Davila-Roman,[33] Gavin Hudson,[90] Xiuqing Guo,[18] Donna K. Arnett,[91] Caroline Hayward,[92] Dhananjay Vaidya,[7] Dennis O. Mook-Kanamori,[15,93] Hemant K. Tiwari,[51] Daniel Levy,[94] Ruth J.F. Loos,[95,96,97] Abbas Dehghan,[34] Paul Elliott,[34] Afshan N. Malik,[98] Robert A. Scott,[14] Diane M. Becker,[7] Mariza de Andrade,[22] Michael A. Province,[1,102] James B. Meigs,[55,99,100,102] Jerome I. Rotter,[18,102] and Kari E. North[4,102,*]

to translate these peptides consisting of 22 transport RNAs (tRNAs) and 2 ribosomal RNAs (rRNAs). In addition, a noncoding sequence, known as the displacement loop (*D-loop*), encompasses the replication origin(s) and promoters for mtDNA.

MT function is also dependent upon many nuclear genes that encode proteins involved in mtDNA transcription, replication, cell apoptosis and mitophagy, nucleotide biosynthesis, metabolism, and iron and calcium homeostasis. Here we evaluate 2,282 candidate nuclear genes that also may contribute to MT function. We refer to these genes as "MT-nDNA candidates," curated for this study based upon protein co-localization within the MT and text mining.[6–8] Unlike mtDNA-encoded proteins that are transcribed and translated within MT, MT-nDNA genes are translated in the cytoplasm and their proteins are transported into MT through porins or complexes such as translocases.

Medicine, Cambridge CB2 0QQ, UK; [15]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden 2333 ZA, the Netherlands; [16]Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas 96020-220, Brazil; [17]MRC Integrative Epidemiology Unit (IEU), Population Health Sciences, University of Bristol, Bristol BS8 2BN, UK; [18]Institute for Translational Genomics and Population Sciences, LABioMed and Department of Pediatrics, at Harbor-UCLA Medical Center, Torrance, CA 90502, USA; [19]Department of Medical Research, Taichung Veterans General Hospital, Taichung 407, Taiwan; [20]Department of Social Work, Tunghai University, Taichung 407, Taiwan; [21]Department of Epidemiology and Biostatistics, University of Illinois at Chicago, Chicago, IL 60612, USA; [22]Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; [23]Department of Genetics, Washington University School of Medicine, St Louis, MO 63110, USA; [24]School of Medicine, Division of Endocrinology, Diabetes and Nutrition, and Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA; [25]Division of Endocrinology, Metabolism and Lipid Research, Washington University School of Medicine, St. Louis, MO 63110, USA; [26]Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; [27]Center for Diabetes Research, Wake Forest School of Medicine, Cincinnati, OH 45206, USA; [28]Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI 53226, USA; [29]The Danish Aging Research Center, University of Southern Denmark, Odense 5000, Denmark; [30]Department of Internal Medicine, Erasmus Medical Center, 3000 CA Rotterdam, the Netherlands; [31]Samford University McWhorter School of Pharmacy, Birmingham, Alabama, Translational Genomics Research Institute (TGen), Phoenix, AZ 35229, USA; [32]Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden 2333 ZA, the Netherlands; [33]Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St Louis, MO 63110, USA; [34]Department of Biostatistics and Epidemiology, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, UK; [35]Department of Hygiene and Epidemiology, University of Ioannina, Ioannina 45110, Greece; [36]Steno Diabetes Center Copenhagen, Copenhagen 2820, Denmark; [37]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; [38]Department of Health Sciences, University of Leicester, Leicester LE1 7RH, UK; [39]Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853, USA; [40]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Centre for Genomic and Experimental Medicine, Medical Genetics Section, University of Edinburgh, Edinburgh EH4 2XU, UK; [41]Department of Internal Medicine, University of Utah, Salt Lake City, UT 84132, USA; [42]Department of Genetic Medicine, Weill Cornell Medicine, PO Box 24144, Doha, Qatar; [43]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL 35294, USA; [44]Biomedical and Translational Informatics, Geisinger Health, Danville, PA 17822, USA; [45]Department of Ophthalmology, Taichung Veterans General Hospital, Taichung 407, Taiwan; [46]Genomics & Pathology Services, Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA; [47]GeneSTAR Research Program, Divisions of Allergy and Clinical Immunology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; [48]Taub Institute for Research on Alzheimer disease and the Aging Brain, Columbia University Medical Center, New York, NY 10032, USA; [49]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald 17475, Germany; [50]University of Minnesota School of Public Health, Division of Epidemiology and Community Health, Minneapolis, MN 55454, USA; [51]Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL 35294, USA; [52]Department of Epidemiology & Population Health, Albert Einstein School of Medicine, Bronx, NY 10461, USA; [53]Department of Endocrinology, Diabetes and Metabolism, Rigshospitalet, Copenhagen University Hospital, 2100 Copenhagen, Denmark; [54]The Danish Diabetes Academy, 5000 Odense, Denmark; [55]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA; [56]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, UK; [57]Institute for Community Medicine, University Medicine Greifswald, 17475 Greifswald, Germany; [58]Department of Cardiology, The Heart Centre, Rigshospitalet, University of Copenhagen, Copenhagen 2100, Denmark; [59]Computer Science and Networking, Wentworth Institute of Technology, Boston, MA 02115, USA;

(Affiliations continued on next page)

The dual genomic origin (mtDNA and MT-nDNA) of MT components and the fact that mtDNA has a much higher mutation rate compared to the nuclear genome[9,10] fuels our hypothesis that MT variants influence individual predisposition to complex diseases.[11,12] Because of heteroplasmy, i.e., the coexistence of mutated and wild-type mtDNA in the same cell, mutations in mtDNA may not result in disease until the copy number of mtDNA carrying the mutation relative to the wild-type mtDNA in the relevant tissue(s) exceeds a threshold.[12] The mtDNA copy number may serve as an indicator of MT function[13] and was recently shown to be associated with overall mortality[14,15] and various metabolic-related diseases, including cardiovascular disease (CVD).[16–18] Studies have also shown that mitochondria have an important influence on multiple cardiovascular disease risk factors. For example, obesity impairs MT biogenesis.[19–21] MT dysfunction has been also implicated in the development of insulin resistance.[22–24] Thus, our study prioritized obesity and insulin resistance pathways to CVD.

In our study we analyzed the dual genomic influence of MT function on seven important metabolic traits (body mass index [BMI], waist-hip ratio [WHR], fasting glucose, fasting insulin, HOMA-B, HOMA-IR, and HbA1c) as major risk factors of cardiovascular disease (CVD) in participants of the CHARGEmtDNA+ 45 participating cohorts. The mtDNA associations were evaluated at both the single nucleotide variant (SNV) and at the gene-based levels for rare alleles (minor allele frequency [MAF] < 1%). The MT-nDNA candidates were studied for function, gene expression regulation, enrichment of pathways, and protein interactions. The overall goal was to illuminate MT-functioning and MT-disease causative variants.

Subjects and METHODS

CHARGE Cohorts

There were 45 participating cohorts with a total of 196,967 individuals (166,423 individuals were of the European [EA], 15,745 of African [AA], 1,567 of Asian [ASA], 11,307 of Hispanic/Latinos [HA], and 1,925 of Brazilian ancestry) who had at least one of the seven traits. CHARGE cohort summary descriptions are available online (see Supplemental Study Descriptions). The number of samples available from all cohorts differed by phenotype (Table S1). The resulting samples were smaller within each cohort when association tests were performed (by sub-setting individuals with genotyping available) to a total of 170,202 for BMI, 155,396 for WHR, 79,906K for fasting glucose, 70,778 for fasting insulin, 69,845 for HOMA-B, 69,926 HOMA-IR, and 101,218 for HbA1c. We use the acronym PA to refer to pan-ancestry. The procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and that proper informed consent was obtained.

Trait Harmonization

A total of seven traits were studied: BMI (kg/m^2), WHR (waist-hip ratio), fasting plasma glucose (mg/dL), fasting insulin (μIUmL),

[60]Genome Technology Access Center, Washington University School of Medicine, St Louis, MO 63110, USA; [61]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [62]Department of Public Health, Section of Epidemiology, Aarhus University, Denmark, Danish Diabetes Academy, Odense University Hospital, 5000 Odense, Denmark; [63]Division of Nephrology and Hypertension, Department of Cardiovascular Diseases, Mayo Clinic, Rochester, MN 55902, USA; [64]Department of Health Science and Technology, Aalborg University Hospital, Aalborg 9220, Denmark; [65]Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung 407, Taiwan; [66]Institute of Medical Technology, National Chung-Hsing University, Taichung 402, Taiwan; [67]School of Medicine, National Defense Medical Center, Taipei 114, Taiwan; [68]School of Medicine, National Yang-Ming University, Taipei 112, Taiwan; [69]Institute of Social and Preventive Medicine (ISPM), University of Bern, 3012 Bern, Switzerland; [70]Department of Medicine, Geriatrics Section, Boston University School of Medicine and Boston Medical Center, Boston, MA 02118, USA; [71]Research Unit for Gynecology and Obstetrics, Department of Clinical Research, University of Southern Denmark, 5000 Odense, Denmark; [72]Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA 15261, USA; [73]Department of Clinical Experimental Research, Rigshospitalet, Copenhagen 2200, Denmark; [74]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2200, Denmark; [75]The Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen 2000, Denmark; [76]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA; [77]Research Centre for Prevention and Health, Glostrup Hospital, Glostrup 2600, Denmark; [78]Department of Public Health, Faculty of Health Sciences, University of Copenhagen, Copenhagen 1014, Denmark; [79]Faculty of Medicine, Aalborg University, Aalborg 9100, Denmark; [80]Novo Nordisk Foundation Center for Basic Metabolic Research (Section of Metabolic Genetics) and Department of Public Health (Section on Epidemiology), Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2200N, Denmark; [81]Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA; [82]Department of Internal Medicine, Section of Endocrinology, Vejle Lillebaelt Hospital, 7100 Vejle, Denmark; [83]Center for Complex Disease Genomics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [84]Department of Clinical Biochemistry, Vejle Hospital, 7100 Vejle, Denmark; [85]Institute of Regional Health Research, University of Southern Denmark, 5000 Odense C, Denmark; [86]Copenhagen Prospective Studies on Asthma in Childhood, Copenhagen University Hospital, Gentofte & Naestved 2820, Denmark; [87]Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark; [88]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA; [89]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA; [90]Wellcome Trust Centre for Mitochondrial Research, Institute of Neuroscience, Newcastle University, Newcastle upon Tyne NE1 3BZ, UK; [91]University of Kentucky, College of Public Health, Lexington, KY 40508, USA; [92]MRC Human Genetics Unit, University of Edinburgh, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh EH4 2XU, UK; [93]Department of Public Health and Primary Care, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands; [94]The Framingham Heart Study, Framingham, MA, USA; The Population Sciences Branch, NHLBI/NIH, Bethesda, MD 20892, USA; [95]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [96]Genetics of Obesity and Related Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [97]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [98]King's College London, Department of Diabetes, School of Life Course, Faculty of Life Sciences and Medicine, London SE1 1NN, UK; [99]Division of General Internal Medicine, Massachusetts General Hospital, Boston 02114, MA, USA; [100]Program in Medical and Population Genetics, Broad Institute, Boston, MA 02142, USA
[101]These authors contributed equally to this work
[102]Co-senior author
*Correspondence: aldikraja@wustl.edu (A.T.K.), kari_north@unc.edu (K.E.N.)
https://doi.org/10.1016/j.ajhg.2018.12.001.

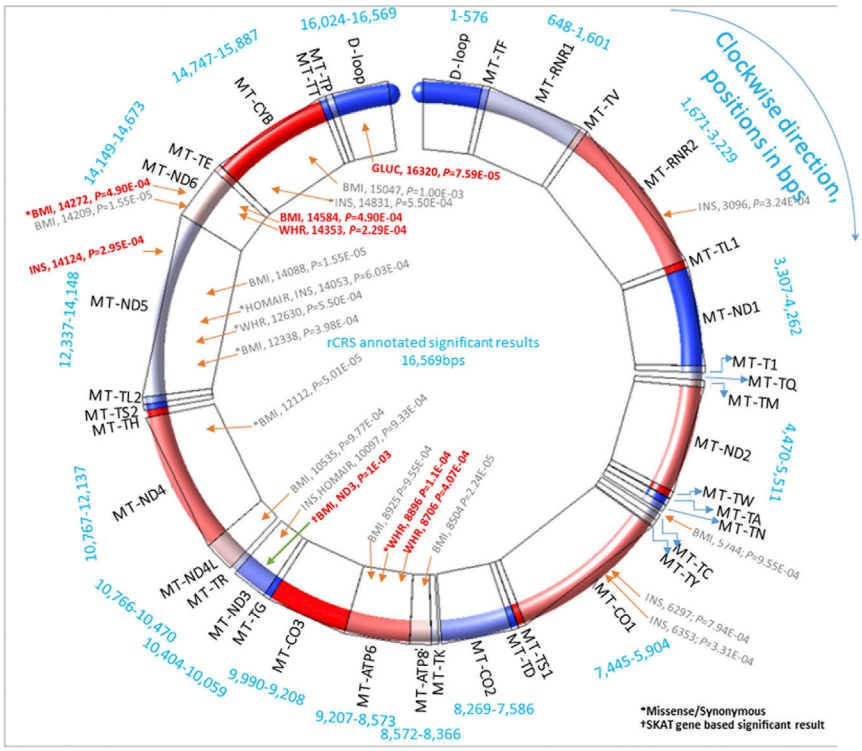


Figure 1. Graphical Presentation of Single mtDNA- and Gene-Based Association Meta-Analysis Results

Significant associations (trait, MT position, and p value) at liberal threshold ($p \leq 1E{-}03$) are annotated with red arrows and gray text for single SNV associations and with green arrow for gene-based association, and text in red color represents SNVs that passed the threshold $p \leq 5E{-}04$. The asterisk (*) marks missense or synonymous variants; a dagger (†) marks a SKAT gene-based significant association. The mtDNA is shown as one single molecule, by merging all genes from heavy and light MT strands. For visibility, gene colors rotate clockwise from first in blue in the *D-loop* to up to six coloring groups finishing in red at *MT-TL1*, and then recycling again from the first blue color at *MT-ND1*. The full information is summarized in Tables 1, 2, 3, and S2.a-S2.b. In blue text are shown bps-positions of 13 MT-coding genes (*MT-ND1*, *MT-ND2*, *MT-CO1*, *MT-CO2*, *MT-ATP8*, *MT-ATP6*, *MT-CO3*, *MT-ND3*, *MT-ND4L*, *MT-ND4*, *MT-ND5*, *MT-ND6*, *MT-CYB*), 2 MT-ribosomal RNAs (MT-RNR1 [12S RNA] and MT-RNR2 [16S RNA]), and 22 transport RNAs (*MT-TF*, *TV*, *TL1*, *T1*, *TQ*, *TM*, *TW*, *TA*, *TN*, *TC*, *TY*, *TS1*, *TD*, *TK*, *TG*, *TR*, *TH*, *TS2*, *TL2*, *TE*, *TT*, *TP*).

HOMA-B, HOMA-IR, and glycated hemoglobin (HbA1c, %) indicative of chronically elevated glucose levels. The homeostasis model assessment was based on plasma levels of fasting glucose and fasting insulin. HOMA-B is an indicator of beta-cell function calculated as HOMA-B = (360 × fasting insulin) / (fasting glucose − 60). HOMA-IR is an indicator of insulin resistance calculated as HOMA-IR = (fasting glucose × fasting insulin) / 405. Glucose and insulin measures were required to have been measured with a minimum fasting time of 8 hr; otherwise, the measure was set to missing. Each study was responsible for ensuring that each trait was normally distributed. Natural-log transformation was implemented for insulin, power transformation was performed when no other solutions were available, and inverse normal transformation was used as a last resort for BMI[25] (details in the Supplemental Material and Methods, section 1). The association test analysis was restricted to non-diabetic participants when studying glycemic traits, by setting values of glucose, insulin, and HbA1c to missing for T2D individuals with HbA1c > 6.5 or with fasting glucose > 126 mg/dL or using T2D medications. The participating cohorts were required to have at least one of the seven studied metabolic traits.

Preparation of Traits for Analyses

All cohorts used the same units for all traits for the raw measures. We observed differences in the raw measures of traits (Table S1). Thus, to harmonize results, all participating cohorts normalized trait distributions and produced standardized residuals, which are unit-less for each trait. Statistical regression via R / SAS programming languages was used to produce residuals fitting the following model:

$$\text{trait} = \text{age} + \text{age}^2 + \text{sex} + \text{PCs} + \text{other study specific covariates (e.g., study center, etc)}$$

Age^2 was used to remove any age-quadratic trend from the response variable, and principal components (PCs) represent one or more genotype principal components, to minimize population substructure. The standardized residuals from this regression were the final response variables for each trait, used in the association tests with imputed dosages of mtDNA.

mtDNA Variant Harmonization

All mtDNA variants from each array (Table S14) were annotated to the nucleotide position according to the rCRS of the human mitochondrial DNA prior to analyses (see Web Resources). The probes used for each microarray were obtained from the manufacturer or dbSNP and aligned to the rCRS using Geneious 8.1.[26] All probes were also submitted through the standard nucleotide basic local alignment search tool (BLAST) to ensure the probes bound with high specificity (≥90% identity) to the mitochondrial genome. In order to limit any potential binding to nuclear DNA segments, probes that bound to nuclear chromosomes with ≥80% matching were excluded from all analyses. The full lists of the validated mtDNA variants are available in the Supplemental Material and Methods, section 2 and Tables S15–S28.

mtDNA Imputation

The preparation of mtDNA data for imputation was done with PLINK.[27,28] When preparing data for imputation, a few heterozygotes that may have existed in the mtDNA genotypes were set to missing via PLINK. The prephasing of mtDNA scaffold haplotypes at the cohort level was done with SHAPEIT2 for the full mtDNA (see Web Resources).[29] SHAPEIT2 helped to have the mtDNA genotypes of a cohort in the right format to be used in the imputation pipeline with IMPUTE2. SHAPEIT2 was also used for QC checks, such as evaluating missingness in markers. In order to combine

analyses across different genotyping platforms, we performed imputation based on MT-1000G (see Supplemental Material and Methods, section 3). Imputations of mtDNA variants were implemented in IMPUTE2 at the cohort level in a window that included the full MT-chromosome (see Web Resources).[30] Recoding of genotype probabilities into dosages was implemented via FCGENE (see Web Resources). Our reference MT-chromosome panel consisted of the 1000G data (PHASE 3, v.5) (see Web Resources). These data were based on sequence freeze and alignments, using only Illumina platform sequences and only sequences with read lengths of 70 bps or greater. There were 3,617 mtDNA markers in 1000G. The 1000G included 2,504 individuals representing 26 sampled populations (the "Cosmopolitan" reference panel).[31] After imputation, cohorts had differing number of SNVs, up to 3,832 SNVs for BMI, 3,829 for WHR, 3,809 for glucose, 3,801 for insulin, 3,801 for HOMA-B, 3,801 for HOMA-IR, and up to 3,744 for HbA1c. Although the imputation quality was good for most of cohorts (see imputation accuracy in Table S14), a great number of the SNVs were filtered for INFO (imputation quality) < 0.30, monomorphic or very rare (MAF < 0.0001) at the cohort level before association tests. Other QC were genotype call per cell had to be $p \geq 0.90$, a marker was dropped if its missing rate was >0.05 and MAC had to be ≥ 5, (which if it was autosomal could have corresponded to MAC ≥ 10). The MAC was calculated for haplotypes as MAF*N. (See Supplemental Material and Methods, section 3 for a more detailed methodology.) After mtDNA imputation, before performing statistical associations, we set any heteroplasmic mutations to missing at the cohort level data. Details of QC via R Package: EasyQC (V10.0)[32] (see Web Resources) and our internal programs and of the imputation steps can be found in the Supplemental Material and Methods, section 3. All filters implemented at the cohort level created non-uniform contributions of all cohorts in the meta-analysis. Many of the mtDNA variants were rare and per trait meta-analyses QQ-plots often showed an underestimation compared to expected association results (Figures S1–S7). The rarity of alleles was accompanied with lower quality of imputation and accordingly many of them were removed before any meta-analysis. Most studies had good imputation quality, given the rarity of mtDNA variants.

Variants MT-14272 and MT-14584 are opposite C/T versus T/C, but in full LD, while they are rare. The typical assumption is that rare variants are not in LD, but that does not have to hold for mtDNA. This is one more observation that the rare association findings in this study represent potential associations, until replicated from other future studies.

mtDNA Association Statistical Analyses

An additive genetic model was applied in association analyses using both self-developed regression models (the linear or linear mixed models written in R programming) and the SKAT (the prepScores() function in seqMeta R package: seqMeta package**:** Meta-Analysis of Region-Based Tests of Rare DNA Variants approaches[33] (see Web Resources). Familial and maternal correlation structures[34] were accounted for in the analysis of family data. Details of statistical association models 1–4 are described in the Supplemental Material and Methods, section 4.

mtDNA Meta-Analyses

Single-variant fixed-effects meta-analyses were conducted with METAL,[35] and gene-based associations were performed in seqMeta (details in the Supplemental Material and Methods, sections 5–7) (Tables 1, 2, 3, S2.a, S2.b, and S13). The average allele frequency from Metal is an average of allele frequencies for a particular marker as contributed from each study and weighted by each study's sample size as follows: $Freq1WeightAve = \sum_{k=1}^{s} w_k f_k / \sum_{k=1}^{s} w_k$, where k represents the indexing of contributing studies, w_k is the sample size (number of individuals) per study, and f_k is the coded allele frequency for a particular marker from a particular study. Our working group conducted an internal permutation test[36] using ARIC study mtDNA data and determined that 49 independent mitochondrial variants represented an estimate of the number of independent genetic effects for mtDNA. Thus, 49 was used as a denominator for producing the Bonferroni corrected p significance threshold $\leq 0.05/49 \leq 1E{-}03$, a threshold for common variants (MAF $\geq 1\%$), but possibly liberal for rare variants (MAF $< 1\%$) (Table 1). Furthermore, we considered $p \leq 0.05/(49 \times 7 \text{ traits}) \leq 1.5E{-}04$ as a conservative threshold when accounting for seven traits tested. We settled for a Bonferroni-threshold $p \leq 0.05/(49 \times 2 \text{ domains}) \leq 5E{-}04$, because the traits within a domain adipose/obesity (BMI and WHR) and glucose metabolism (glucose, insulin, HOMA-B, HOMA-IR, and HbA1c) are correlated and represent two domains. Because mtDNA is a small genome, the distributions of mtDNA QQ-plots with the existing samples do not always behave similarly to those observed for larger nuclear GWASs. While mtDNA variants MAF $\geq$ 1% formed relatively good QQ-plot distributions, the rare variants (MAF $< 1\%$) were sometimes distributed with some deviation in the start (bottom-left) of the QQ-plot, as a result of meta-results in very rare alleles. Thus, the quality control of mtDNA QQ-plots were implemented by filtering at different rare MAF levels. We concluded that with the existing mtDNA meta-analysis results, QQ-plots with MAF $> 0.8\%$ were acceptable. Thus, out of 23 SNVs nominally ($p \leq 1E{-}03$) associated with 6 metabolic traits (Figure 1), by adding the filter of MAC ≥ 5 for each cohort, then 7 variants passed $p \leq 5E{-}04$ Bonferroni threshold.

For the gene-based analysis, we used Burden tests which combine the contributions of rare genetic variants within a gene region. Such tests assume similar directionality and effect sizes for each variant.[33] In contrast, the sequence kernel association test (SKAT) for unrelated or family-based studies is an efficient regression-based approach for rare genetic variant analysis.[33,37,38] The meta-analyses were performed using seqMeta (see Web Resources). The gene-based Bonferroni p value was calculated as $p \leq 0.05/$ (37 mtDNA genes) $\leq 1E{-}03$.

Identification of MT-nDNA Candidate Genes

We established a list of 2,282 MT-nDNA candidate genes from three sources: (1) Mito Carta 2.0,[6,39] (2) Literature Lab (Acumenta Biotech), and (3) MitoMiner (4.0)[7] (see also Supplemental Material and Methods, section 7). We used the two separate sets of human genes and mouse ortholog genes from MitoCarta. We used Literature Lab to perform a literature search and identified 36 terms based on MeSH mitochondria. From each of the 36 terms (Table S29), we selected only the upper quartile list of genes from a Log Probability Function - scoring distribution. Each term was tested for association in overlapping with genes in pathway analysis (Figure S15). We conditioned this list with only genes from human nomenclature and accepted only genes that had more than 15 abstracts cited per selected gene (see Supplemental Material and Methods section 7).

Using MitoMiner we identified additional MT-nDNA candidate genes. They were filtered with a MT-MitoMiner index ≥ 0.70,

Table 1. mtDNA Variants Associated with BMI, WHR, Glucose, Insulin, HOMA-B, HOMA-IR, and HbA1c METAL Meta-Analysis Single-Variant Results

No	Pos	rsID	Gene	Annotation	Trait	A1/2	Freq1 WeightAve	FreqSE	MinFreq	MaxFreq	MAF	MAC	β(SE)	p	Dir	Het-p	N
Meta-analysis Level Results Selected with MAF of Weighted Allele Average Frequency > 1% and Passing Meta-p-Threshold ≤ 5E−04																	
1	8706	–	MT-ATP6	–	WHR	A/G	0.9676	0.0295	0.9135	0.9974	0.0324	834	−0.13 (0.04)	4.1E−04	----+	9.44E−02	25,748
2	16320	rs62581338	D-loop	–	GLUC	T/C	0.0158	0.0954	0.0028	0.2439	0.0158	301	−0.21 (0.05)	7.6E−05	--+--	1.77E−01	19,046
Meta-analysis Level Results Selected with MAF of Weighted Allele Average Frequency < 1% and Passing Meta-p-Threshold ≤ 5E−04																	
1	8896	rs202120082	MT-ATP6	missense	WHR	A/G	0.0038	0.0045	0.0003	0.0121	0.0038	134	0.30 (0.08)	1.12E−04	++++	8.80E−01	34,959
2	14124	–	MT-ND5	–	INS	T/C	0.0035	0.0029	0.0022	0.0087	0.0035	21	0.57 (0.16)	2.95E−04	+++	1.62E−01	6,035
3	14272	rs2853814	MT-ND6	missense	BMI	T/C	0.0012	0.0007	0.0009	0.0022	0.0012	17	0.84 (0.24)	4.90E−04	++	6.49E−01	13,636
4	14353	–	MT-ND6	–	WHR	T/C	0.9916	0.0032	0.9880	0.9978	0.0084	120	−0.33 (0.09)	2.29E−04	--	8.00E−01	14,315
5	14584	–	MT-ND6	–	BMI	T/C	0.9988	0.0007	0.9978	0.9991	0.0012	17	−0.84 (0.24)	4.90E−04	--	6.49E−01	13,636

Abbreviations and definitions: No, order number; Pos, MT position in bps; rsID, rsID name from NCBI dbSNP database when available; Gene, gene name or region; Annotation, role of the variants when available; Trait, one or more of seven traits studied; A1/2, the coded and non-coded alleles; Freq1 WeightAve, a weighted sample size average of allele frequency; FreqSE, standard error of allele frequency; MINFreq, a minimum allele frequency for contributing cohorts; MAXFreq, a maximum allele frequency for contributing cohorts; MAF, minor allele frequency; MAC, minor allele count, calculated as MAF × N; β(SE), beta coefficient and the corresponding standard error; p value, from single variant regression analysis; Dir, direction sign of contributing cohort's beta; Het-p, heterogeneity p value test from METAL; N, individuals' sample contributing in a particular marker meta-analysis (all results are of Pan-Ancestry).

and by selecting the terms "Known mitochondrial" and "Predicted mitochondrial." In MitoMiner, we kept only genes that were present in human nomenclature. MitoMiner included also the mitochondrial originated genes, which were later removed to keep our MT-nDNA list only of nuclear origin. Finally, three additional genes were added from a publication on MT-defects associated with β-cell dysfunction in a T2D mouse.[40]

MT-nDNA Candidate Genes Analysis

As a result of the above work, a list of 2,282 MT-nDNA candidate gene labels (Table S30) were used to identify any significant results from 20 GWAS papers full summary results for seven metabolic traits, representing 31 datasets (Tables 4, S4, and S6). For the adipose traits, the GWAS publication full summary results used were for BMI[41–50] and for WHR,[43,44,47–52] and the summary results data were retrieved from the GIANT Consortium repositories (see Web Resources). For glucose metabolism we used the summary results of glucose,[53–56] insulin,[54–59] HOMA-B,[54] and HOMA-IR.[54] These summary results data were retrieved from MAGIC Consortium archives (see Web Resources). For HbA1c we used two resources.[60,61] The published GWASs have large sample sizes, with a BMI study having a maximum of 339K individuals, WHR having 320K, glucose and insulin having 52K and 45K, respectively, HOMA-B and -IR having 46K, and HbA1c having 160K. Consequently, we obtained results for 109 MT-nDNA candidate genes, accompanied by 588 sentinel significant SNVs (one best per gene and trait combination, out of seven traits, Tables 4, S6, and S9).

To identify the 588 SNVs, a pre-specified selection process was followed. First, we downloaded the latest dbSNP of NCBI reference data (batch 150), and we assigned each SNV to a gene (pseudogenes excluded). The intergenic variants were assigned to the closest gene, up to half the distance between two genes. Then, we merged the 31 full-GWAS sets to the annotated dbSNP, thus each GWAS-SNP was assigned to a gene. Each summary result was merged with MT-nDNA gene labels (Table S30). The corresponding significant SNVs results, ones with the smallest p value, per gene and per trait, were accepted in the final list ($p < 0.05/(2{,}282 \times 7) = 3.13E{-}06$). After this selection, we also performed an analysis based on 1000 Genomes to identify the number of independent SNVs within a gene.[36,62] The analysis produced a mean of independent SNVs per gene of 59, median of 38, and with a maximum of two outliers 542 (*WWOX* [MIM: 605131]) and 499 (*FHIT* [MIM: 601153]). If it was a gene-based test, then a conservative threshold $p \leq 0.05/(542 \times 7 \text{ traits}) \leq 1.3E{-}05$ could have been used, which is larger than the threshold we used, $p \leq 1E{-}06$). Furthermore, to compare whether our MT-nDNA genes pass the genome-wide threshold of $p \leq 5E{-}08$, generally used in GWAS publications, we merged our gene data with all possible reported SNVs from the GWAS-Catalog (accessed 01.27.2018, Table S12 and presented findings in the Results and Discussion paragraphs).

Annotation, Enrichment Analysis, and Gene Expression and Regulation

The mtDNA as well as MT-nDNA significant variants were annotated to NCBI dbSNP build 150 (HG38); genes and their protein biological functions were annotated to NCBI Entrez Gene, GeneCards and UniProtKB; enrichment analyses were performed with MetaCore and Literature Lab; pathways with KEGG and Reactome; gene expression and regulation were assessed using

Table 2. Results of mtDNA Gene-Based Meta-analysis SKAT T1 and T5 Test ($p \leq 0.01$)

Trait	Ancestry	T	Gene	p	Qmeta	cMAF	nSNVs
HOMA-B	PA	0.01	MT-TF	5.0E−03	4604374	0.022	5
HOMA-B	PA	0.05	MT-TV	8.0E−03	3341460	0.078	6
HbA1c	EA	0.05	MT-TG	9.0E−03	13674145	0.064	13
BMI	PA	0.05	MT-TQ	1.0E−02	21201188	0.075	12
HOMA-B	PA	0.05	MT-RNR1	1.0E−02	69787232	1.323	119

Abbreviations and definitions: Trait, the trait used for a specific test; PA, Pan ancestry; EA, European ancestry; T, MAF-value threshold for selecting SNVs to be included in the gene-based association test; Gene, gene name from mtDNA; p, p value from the SKAT test; Qmeta, the SKAT Q statistics; $Q = \sum_j w_j^2 U_j^2$, where w is a weight for SNV_j and U_j is associated score statistics; cMAF, cumulative minor allele frequency; nSNVs, the number of SNVs used in the gene-based meta-analysis.

HaploReg, RegulomeDB, GTEx, and Human Protein Atlas; and protein interactions were assessed using STRING and NCBI summary of interactions from other databases (see Web Resources), with references to databases BIND, BioGRID, EcoCyc, HIV-1-human protein interaction data, and HPRD. Specifically, for GTEx gene expression eQTL analysis, the eSNVs were considered significant when the eSNV had a GTEx $p < 1E{-}07$ and was in high LD ($r^2 \geq 0.80$) with the best eSNV of the target gene. Depending on $r^2 \geq 0.80$ to 1, we called them similar to "lead" or "lead" eSNVs. Otherwise, if the LD r^2 was < 0.02 to the target gene's best eSNV, we called them "secondary" eSNVs (Table S8). Details of the resources and the corresponding references are provided in the Supplemental Material and Methods, sections 8 and 9. We have cited throughout the manuscript the corresponding gene MIM number from the Online Mendelian Inheritance in Man (see Web Resources).

Results

We evaluated the associations of mtDNA variants with seven key metabolic traits in meta-analyses of 45 cohorts (with up to N∼170,202). For details on the harmonization of the phenotypes and genotypes, see Material and Methods and Supplemental Material and Methods, sections 1–6. The Supplemental Study Descriptions and Table S1 with BMI mean and standard deviation values give a depiction of each contributing cohort in this study.

mtDNA Single-Variants Associations

Seven SNVs, two variants with average weighted MAF > 1% (Tables 1 and S2.a and Figures S1–S7) and five with MAF < 1% (Tables 1 and S2.b) displayed statistically significant evidence of association with six metabolic traits (Bonferroni threshold $p \leq 5E{-}04$, see in Material and Methods: mtDNA Meta-Analyses). MT-8706 in *MT-ATP6* (MIM: 516060) (see Web Resources), associated with WHR (with sample weighted average of cohorts' MAF = 3.24%, $p = 4.1E{-}04$) and MT-16320 (rs62581338) of the *D-loop* (with sample weighted average of cohorts' MAF = 1.58%, $p = 7.6E{-}05$) associated with fasting glucose. The five rare variants were MT-8896 (rs202120082, missense) in *MT-ATP6* associated with WHR, MT-14124 in *MT-ND5* (MIM: 516005) associated with fasting plasma insulin, MT-14272 in *MT-ND6* (rs2853814 [MIM: 516006], missense) associated with BMI, MT-14353 in *MT-ND6* associated with WHR, and MT-14584 in *ND6* associated with BMI.

The evidence of association by cohort is reported in Tables S2.a, S2.b, and S3. Typically, the cohort with a large sample size displayed the strongest evidence of association. For example, the HCHS/SOL study contributed disproportionately to several mtDNA-trait associations (see in the Supplemental Study Descriptions, CHARGEmtDNA+ Study Description). Even further, for the *MT-ATP6* (position 8706) association with WHR, the strongest association in HCHS/SOL was in those of Central/South American background (so more Native American ancestry as compared to African American ancestry, Tables S2.a and S3).

mtDNA Rare Variants Gene-Based Associations

The mitochondrial rare variants for gene-based analysis were mapped to the start and end positions of each gene from the NCBI Reference Sequence GenBank: NC_012920.1 (see Web Resources). The rare variant gene-based meta-analysis using SKAT did not yield any significant associations. In contrast, the Burden test yielded a significant association between rare variants in *MT-ND3* and BMI ($p = 1E{-}03$, $T < 0.05$ including 82 SNVs). A forest plot representing the 82 SNVs and the overall MT-ND3-overall meta is shown in Figure S17. Several gene-HOMA-B, HbA1c, and BMI associations were suggestively significant employing both the SKAT and Burden approaches (MT-RNAs *MT-TF* [MIM: 590070], *MT-TV* [MIM: 590105], *MT-TG* [MIM: 590035], *MT-TQ* [MIM: 590030], and *MT-RNR1* [MIM: 561000]; $p \leq 1E{-}02$; Tables 2 and 3). A total of 131 (T5-test) and 123 (T1-test) low-frequency and/or rare variants contributed to T5-test and T1-test, respectively. We used two gene-based approaches, burden test and sequence kernel association test (SKAT), to evaluate the association between a gene and a phenotype. Burden tests access the cumulative effects of multiple variants by assuming that all variants have the same directionality in a genetic region. SKAT tests use a score-based variance component framework without assuming that all variants have the same directionality. Although the two methods are different, they all evaluate aggregate effects of multiple low-frequency and/or rare variants

Table 3. Results of mtDNA Gene-Based Burden T1 and T5 Meta-Analysis Test ($p \leq 0.01$)

Trait	Ancestry	T	Gene	p	Beta	SE	cMAFUsed	nSNVs
BMI	PA	0.05	MT-ND3	1.0E−03	0.007	0.002	0.290	82
BMI	PA	0.05	MT-TQ	2.0E−03	0.020	0.006	0.075	12
BMI	PA	0.05	MT-CO2	5.0E−03	0.003	0.001	0.491	155
HOMAB	PA	0.01	MT-TF	5.0E−03	0.038	0.014	0.022	5
HOMAB	PA	0.01	MT-TV	6.0E−03	0.088	0.032	0.009	2
HOMAB	PA	0.05	MT-RNR1	7.0E−03	0.002	0.001	1.323	119
HOMAIR	EA	0.05	MT-TG	7.0E−03	0.027	0.010	0.061	13
BMI	EA	0.05	MT-ND3	7.0E−03	0.006	0.002	0.194	81
HOMAB	PA	0.05	MT-TV	8.0E−03	0.031	0.011	0.078	6
BMI	EA	0.05	MT-TQ	8.0E−03	0.018	0.007	0.064	12
HOMAB	EA	0.05	MT-RNR1	9.0E−03	0.002	0.001	1.287	120
HOMAB	EA	0.05	MT-TF	9.0E−03	0.018	0.007	0.127	10
HOMAB	EA	0.05	MT-TV	9.0E−03	0.030	0.011	0.079	6
HOMAB	EA	0.01	MT-RNR1	1.0E−02	0.003	0.001	0.292	61

Abbreviations and definitions: Trait, the trait used for a specific test; PA, Pan ancestry; EA, European ancestry; T, MAF-value threshold for selecting SNVs to be included in the gene-based-burden association test; Gene, gene name from mtDNA; p, p value from the gene-based burden test; the score test of the weighted sum of genotypes has the form of statistic, $T = \sum_j w_j U_j$, where w is a weight for SNV_j and U_j is the score statistic for SNV_j; Beta and SE are result of regressing the trait on a weighted sum of genotypes; cMAFUsed, cumulative minor allele frequency; nSNVs, the number of SNVs used in the gene-based burden meta-analysis.

with a phenotype. A significant gene-test reflects an aggregate effect of multiple variants in the same area. If most variants in *ND6* are not associated with the phenotype, the gene-based test might not be significant. Although statistical significance was not achieved, the association of MT-ND6 with BMI was suggestive in the burden (p = 0.06 [T5] and 0.04 [T1]) and SKAT (p = 0.15 [T5] and 0.08 [T1] tests).

Identification of MT-nDNA Candidate Variants Associated with Metabolic Traits

We identified 2,282 MT-nDNA candidate genes (see Material and Methods) and assessed their association from 20 GWASs with 31 datasets (Table S4). From the MT-nDNA candidate list, 109 genes reached statistical significance following correction for multiple testing (Bonferroni p < 1E−06) of which 46 were associated with BMI, 2 with extreme obesity, 26 with WHR, 18 with glucose, 7 with insulin, 1 with HOMA-B, 1 with HOMA-IR, and 20 with HbA1c, totaling 121 associations (Table 4). The use of additional SNVs belonging to the same 109 MT-nDNA genes, but now sourced from the GWAS catalog for the same 7 traits or any other traits, indicated that 84% (27% improvement) of MT-nDNA genes passed p ≤ 5E−08, while 16% remained at p ≤ 1E−06 without passing the p ≤ 5E−08 threshold. Of the MT-nDNA associations, *GCK* (MIM: 138079), *MRPL33* (MIM: 610059), *PPARG* (MIM: 601487), *SLC2A2* (MIM: 138160), *AMBRA1* (MIM: 611359), *NR1H3* (MIM: 602423), *MTCH2* (MIM: 613221), *IGF1* (MIM: 147440), and *BCL2* (MIM: 151430) were associated with more than one trait (Figures 2, 3, S8, and S9). For example, the well-known *GCK* is a member of hexokinases that phosphorylates glucose to produce glucose-6-phosphate, the first step in most glucose metabolism pathways, and has pleiotropic effects (see Discussion).

Enrichment Analysis of MT-nDNA Candidate Genes

The 109 MT-nDNA selected genes are candidates for MT function based on their protein localization in mitochondria, as well as from mining the published literature (Table S5). We used Literature Lab and MetaCore software and the corresponding databases to perform enrichment analyses, which provided information for MT-nDNA gene-label relations with terms, pathways, diseases, gene ontology processes, and clustering (see Supplemental Material and Methods, section 8 and Figure S10). When comparing the 109 MT-nDNA candidate genes versus the remaining 2,173 (2,282 − 109), the 109 set showed enrichment (p = 1.7E−12) for "Signal transduction, Neuropeptide signaling pathways," which included, among others, *POMC* (MIM: 176830) and *MC4R* (MIM: 155541), while no such enrichment was found for the 2,173 set. Out of 109 MT-nDNA candidate genes, 21 of the significantly associated genes were functionally related with cholesterol (Table S6 and Figure S11), 13 with glucose and insulin, and 5 with adipose/obesity. Of the 109 MT-nDNA genes, 13 associated with the "Type 2 diabetes"[54,61,63] term (Table S7) while 18 were associated with "Cardiovascular disease"[64–77] (Table S7). (For space limitation, a detailed enrichment analysis is provided in the Supplemental Material and Methods, section 8.) These

Table 4. GWAS Findings for Seven Traits (BMI, WHR, Glucose, Insulin, HOMA-B, HOMA-IR, HbA1c) for 109 MT-nDNA Candidate Genes

No	SNV	Chrom	Position	Role	Gene	Full Gene Name (Cytogenetic Position)	BMI p	OBESITY p	WHR p	GLUC p	INS p	HOMA-B p	HOMAIR p	HbA1c p
1	rs622798	1	45549599	upstream-2 kb	AKR1A1	Aldo-Keto Reductase Family 1 Member A1 (1p34.1)	8.56E−07	–	–	–	–	–	–	–
2	rs1280316	1	66843725	Intron	WDR78	WD repeat domain 78 (1p31.3)	6.81E−07	–	–	–	–	–	–	–
3	rs1093013	1	75634658	Intron	SLC44A5	Solute carrier family 44 member 5 (1p31.1)	–	–	4.86E−17	–	–	–	–	–
4	rs6428792	1	119114244	Intron	WARS2	Tryptophanyl TRNA Synthetase 2, Mitochondrial (1p12)	–	–	7.95E−18	–	–	–	–	–
5	rs2301453	1	172389027	Intron	DNM3	Dynamin 3 (1q24.3)	–	–	4.38E−17	–	–	–	–	–
6	rs4844390	1	207761504	Intron	CD46	CD46 Molecule, Complement Regulatory Protein (1q32.2)	–	–	–	–	–	–	–	6.90E−07
7	rs11118296	1	219408638	Intron	LYPLAL1	Lysophospholipase Like 1 (1q41)	–	–	3.00E−07	–	–	–	–	–
8	rs6713865	2	23676937	Intron	KLHL29	Kelch like family member 29 (2p24.1)	–	–	–	–	–	–	–	4.39E−13
9	rs934778	2	25166355	Intron	POMC	Proopiomelanocortin (2p23.3)	7.15E−07	–	–	–	–	–	–	
10	rs1107238	2	26235376	Intron	HADHA	Hydroxyacyl-CoA Dehydrogenase/3-Ketoacyl-CoA Thiolase/Enoyl-CoA Hydratase (Trifunctional Protein), Alpha Subunit (2p23.3)	8.86E−07	–	–	–	–	–	–	
11	rs13404446	2	27296386	intron	TRIM54	Tripartite Motif Containing 54 (2p23.3)	–	–	–	3.09E−09	–	–	–	–
12	rs4665965	2	27313513	intron	MPV17	MPV17, Mitochondrial Inner Membrane Protein (2p23.3)	–	–	–	1.83E−09	–	–	–	–
13	rs3736594	2	27772914	intron	MRPL33	Mitochondrial Ribosomal Protein L33 (2p23.2)	–	–	–	*3.02E−13*	*4.67E−07*	–	–	–
14	rs4346434	2	43992607	intron	LRPPRC	Leucine Rich Pentatricopeptide Repeat Containing (2p21)	–	–	8.52E−08	–	–	–	–	–
15	rs16843390	2	209655109	intron	MAP2	Microtubule Associated Protein 2 (2q34)	–	–	–	3.32E−08	–	–	–	–
16	rs715	2	210678331	utr-3-prime	CPS1	Carbamoyl-Phosphate Synthase 1 (2q34)	5.81E−07	–	–	–	–	–	–	–
17	rs933994	2	218785893	intron	CYP27A1	Cytochrome P450 Family 27 Subfamily A Member 1 (2q35)	8.65E−07	–	–	–	–	–	–	–
18	rs17036328	3	12348985	intron	PPARG	Peroxisome Proliferator Activated Receptor Gamma (3p25.2)	*4.27E−07*	–	–	–	*3.59E−12*	–	–	–

(*Continued on next page*)

Table 4. ***Continued***

No	SNV	Chrom	Position	Role	Gene	Full Gene Name (Cytogenetic Position)	BMI p	OBESITY p	WHR p	GLUC p	INS p	HOMA-B p	HOMAIR p	HbA1c p
19	rs3729931	3	12585017	intron	RAF1	Raf-1 Proto-Oncogene, Serine/Threonine Kinase (3p25.2)	–	–	3.60E−10	–	–	–	–	–
20	rs11715915	3	49417897	nc-transcript	AMT	Aminomethyltransferase (3p21.31)	–	–	–	4.90E−08	–	–	–	–
21	rs12489828	3	52532998	intron	NT5DC2	5′-Nucleotidase Domain Containing 2 (3p21.1)	–	–	2.60E−10	–	–	–	–	–
22	rs2365389	3	61250788	intron	FHIT	Fragile Histidine Triad (3p14.2)	3.75E−15	–	–	–	–	–	–	–
23	rs332375	3	66377079	intron	SLC25A26	Solute Carrier Family 25 Member 26 (3p14.1)	–	–	7.33E−07	–	–	–	–	–
24	rs1735536	3	128377770	intron	EEFSEC	Eukaryotic Elongation Factor, Selenocysteine-TRNA Specific (3q21.3)	2.95E−10	–	–	–	–	–	–	–
25	rs9844666	3	136255374	intron	PCCB	Propionyl-CoA Carboxylase Beta Subunit (3q22.3)	7.22E−07	–	–	–	–	–	–	–
26	rs11924648	3	171000207	intron	SLC2A2	Solute Carrier Family 2 Member 2 (3q26.2)	–	–	–	*1.02E−17*	–	–	–	*4.05E−09*
27	rs10012946	4	6291623	intron	WFS1	Wolframin ER Transmembrane Glycoprotein (4p16.1)	–	–	–	4.17E−07	–	–	–	–
28	rs10518406	4	122841742	intron	FGF2	Fibroblast Growth Factor 2 (4q28.1)	6.50E−07	–	–	–	–	–	–	–
29	rs1458758	4	122914724	intron	NUDT6	Nudix Hydrolase 6 (4q28.1)	–	–	6.22E−08	–	–	–	–	–
30	rs303084	4	123145793	intron	SPATA5	Spermatogenesis associated 5 (4q28.1)	–	–	3.40E−07	–	–	–	–	–
31	rs12654264	5	75352778	intron	HMGCR	3-Hydroxy-3-Methylglutaryl-CoA Reductase (5q13.3)	1.81E−08	–	–	–	–	–	–	–
32	rs10478424	5	119453325	intron	HSD17B4	Hydroxysteroid 17-Beta Dehydrogenase 4 (5q23.1)	–	–	1.40E−07	–	–	–	–	–
33	rs2881156	5	135812973	intron	SLC25A48	Solute carrier family 25 member 48 (5q31.1)	–	8.10E−07	–	–	–	–	–	–
34	rs3828870	6	16743066	intron	ATXN1	Ataxin 1 (6p22.3)	–	–	–	–	–	–	–	5.52E−07
35	rs1800562	6	26092913	intron	HFE	Hemochromatosis (6p22.2)	–	–	–	–	–	–	–	4.67E−28
36	rs1800629	6	31575254	upstream-2KB	TNF	Tumor necrosis factor (6p21.33)	–	–	7.30E−07	–	–	–	–	–
37	rs6457796	6	34860776	intron	UHRF1BP1	UHRF1 binding protein 1 (6p21.31)	1.15E−09	–	–	–	–	–	–	–

(*Continued on next page*)

Table 4. ***Continued***

No	SNV	Chrom	Position	Role	Gene	Full Gene Name (Cytogenetic Position)	BMI p	OBESITY p	WHR p	GLUC p	INS p	HOMA-B p	HOMAIR p	HbA1c p
38	rs10434	6	43785475	utr-3-prime	VEGFA	Vascular endothelial growth factor A (6p21.1)	–	–	8.80E−07	–	–	–	–	–
39	rs1049354	6	88143732	utr-3-prime	CNR1	Cannabinoid receptor 1 (6q15)	9.57E−07	–	–	–	–	–	–	–
40	rs9400239	6	108656460	intron	FOXO3	Forkhead box O3 (6q21)	1.61E−08	–	–	–	–	–	–	–
41	rs1273733	6	121131419	downstream-500B	TBC1D32	TBC1 domain family member 32 (6q22.31)	3.96E−12	–	–	–	–	–	–	–
42	rs1049349	6	121449496	utr-3-prime	GJA1	Gap junction protein alpha 1; synonymous: CX43 (6q22.31)	4.18E−15	–	–	–	–	–	–	–
43	rs1293954	6	151669826	intron	ESR1	Estrogen receptor 1 (6q25.1-q25.2)	4.41E−09	–	–	–	–	–	–	–
44	rs1203576	7	40808233	intron	SUGCT	*C7orf10* (Succinyl-CoA:Glutarate-CoA Transferase, 7p14.1)	1.48E−10	–	–	–	–	–	–	–
45	rs2908289	7	44184343	intron	GCK	Glucokinase (7p13)	–	–	–	*3.32E−88*	–	*6.00E−09*	–	*2.24E−19*
46	rs1088867	7	44705214	intron	OGDH	Oxoglutarate dehydrogenase (7p13)	7.90E−08	–	–	–	–	–	–	–
47	rs16892421	8	106499705	intron	OXR1	Oxidation resistance 1 (8q23.1)	–	–	6.94E−07	–	–	–	–	–
48	rs7835803	8	120030213	intron	DEPTOR	DEP domain containing MTOR interacting protein (8q24.12)	–	–	–	4.97E−07	–	–	–	–
49	rs2777795	9	104910084	intron	ABCA1	ATP binding cassette subfamily A member 1 (9q31.1)	–	–	3.13E−08	–	–	–	–	–
50	rs7023913	9	128255683	downstream-500B	DNM1	Dynamin 1 (9q34.11)	–	7.30E−07	–	–	–	–	–	–
51	rs3829109	9	136362314	intron	DNLZ	DNL-Type Zinc Finger (9q34.3)	–	–	–	1.13E−10	–	–	–	–
52	rs1244497	10	7838019	intron	TAF3	TATA-box binding protein associated factor 3 (10p14)	1.84E−11	–	–	–	–	–	–	–
53	rs5030913	10	69246375	intron	HKDC1	Hexokinase domain containing 1 (10q22.1)	–	–	–	–	–	–	–	3.56E−13
54	rs4745982	10	69330087	intron	HK1	Hexokinase 1 (10q22.1)	–	–	–	–	–	–	–	2.87E−65
55	rs7899106	10	85651147	intron	GRID1	Glutamate Ionotropic Receptor Delta Type Subunit 1 (10q23.2)	2.96E−08	–	–	–	–	–	–	–
56	rs7917772	10	102727686	intron	SFXN2	Sideroflexin 2 (10q24.32)	–	–	1.45E−09	–	–	–	–	–
57	rs1004467	10	102834750	intron	CYP17A1	Cytochrome P450 Family 17 Subfamily A Member 1 (10q24.32)	1.18E−07	–	–	–	–	–	–	–

(*Continued on next page*)

Table 4. ***Continued***

No	SNV	Chrom	Position	Role	Gene	Full Gene Name (Cytogenetic Position)	BMI p	OBESITY p	WHR p	GLUC p	INS p	HOMA-B p	HOMAIR p	HbA1c p
58	rs3740390	10	102878723	intron	AS3MT	Arsenite Methyltransferase (10q24.32)	4.82E−08	–	–	–	–	–	–	–
59	rs4758633	11	219538	intron	SIRT3	Sirtuin 3 (11p15.5)	–	–	–	–	–	–	–	3.44E−10
60	rs757110	11	17396930	missense	ABCC8	ATP Binding Cassette Subfamily C Member 8 (11p15.1)	4.23E−07	–	–	–	–	–	–	–
61	rs10767664	11	27704439	intron	BDNF	Brain Derived Neurotrophic Factor (11p14.1)	5.53E−13	–	–	–	–	–	–	–
62	rs11038913	11	46538180	intron	AMBRA1	Autophagy and beclin 1 regulator 1 (11p11.2)	–	–	–	*4.29E−08*	*4.91E−18*	–	–	–
63	rs11039149	11	47255124	intron	NR1H3	Nuclear Receptor Subfamily 1 Group H Member 3 (11p11.2)	–	–	–	*1.26E−12*	*4.13E−45*	–	–	–
64	rs7118178	11	47637583	intron	MTCH2	Mitochondrial carrier 2 (11p11.2)	*5.12E−08*	–	–	*3.84E−14*	*2.16E−29*	–	–	–
65	rs4246215	11	61796827	utr-3-prime	FEN1	Flap Structure-Specific Endonuclease 1 (11p12.2)	–	–	–	4.46E−11	–	–	–	–
66	rs174556	11	61813163	intron	FADS1	Fatty acid desaturase 1 (11q12.2)	–	–	–	7.82E−18	–	–	–	–
67	rs7943191	11	62561079	intron	EEF1G	Eukaryotic translation elongation factor 1 gamma (11q12.3)	–	–	4.36E−08	–	–	–	–	–
68	rs11231150	11	62584330	intron	TUT1	Terminal uridylyl transferase 1, U6 snRNA-specific (11q12.3)	–	–	5.20E−08	–	–	–	–	–
69	rs1017639	11	68831066	intron	CPT1A	Carnitine Palmitoyltransferase 1A (11q13.3)	4.96E−10	–	–	–	–	–	–	–
70	rs1296252	11	83273268	intron	CCDC90B	Coiled-Coil Domain Containing 90B (11q13.3)	7.60E−08	–	–	–	–	–	–	–
71	rs2110073	12	6966719	intron	PHB2	Prohibitin 2 (12p13.31)	–	–	–	–	–	–	–	4.44E−08
72	rs7311050	12	7013532	intron	LPCAT3	Lysophosphatidylcholine Acyltransferase 3 (12p13.31)	–	–	–	–	–	–	–	8.60E−08
73	rs1049380	12	26336611	downstream-500B	ITPR2	Inositol 1,4,5-Trisphosphate Receptor Type 2 (12p11.23)	–	–	1.42E−13	–	–	–	–	–
74	rs2408955	12	48105348	upstream-2KB	PFKM	Phosphofructokinase, muscle (12q13.11)	–	–	–	–	–	–	–	1.42E−15
75	rs35767	12	102481791	missense	IGF1	Insulin Like Growth Factor 1 (12q23.2)	–	–	–	–	*7.27E−08*	–	*7.57E−08*	–
76	rs4766578	12	111466567	intron	ATXN2	Ataxin 2 (12q24.12)	2.85E−07	–	–	–	–	–	–	1.84E−07
77	rs9581856	13	27451478	upstream-2KB	MTIF3	Mitochondrial Translational Initiation Factor 3 (13q12.2)	1.03E−08	–	–	–	–	–	–	–

(*Continued on next page*)

Table 4. ***Continued***

No	SNV	Chrom	Position	Role	Gene	Full Gene Name (Cytogenetic Position)	BMI p	OBESITY p	WHR p	GLUC p	INS p	HOMA-B p	HOMAIR p	HbA1c p
78	rs1124607	13	27921083	intron	PDX1	Pancreatic And Duodenal Homeobox 1 (13q12.2)	–	–	–	4.49E−07	–	–	–	–
79	rs1325363	13	33192439	intron	STARD13	StAR related lipid transfer domain containing 13 (13q13.1-q13.2)	7.25E−08	–	–	–	–	–	–	–
80	rs1078892	13	40563883	intron	FOXO1	Forkhead Box O1 (13q14.11)	5.11E−08	–	–	–	–	–	–	–
81	rs7143963	14	102838088	intron	TRAF3	TNF Receptor Associated Factor 3	2.82E−07	–	–	–	–	–	–	–
82	rs12908437	15	98744146	intron	IGF1R	Insulin Like Growth Factor 1 Receptor	–	–	–	6.32E−07	–	–	–	–
83	rs740862	16	3639677	intron	DNASE1	Deoxyribonuclease 1	3.21E−07	–	–	–	–	–	–	–
84	rs151181	16	28479196	intron	CLN3	CLN3, Battenin	2.10E−07	–	–	–	–	–	–	–
85	rs8055138	16	28880144	intron	ATP2A1	ATPase Sarcoplasmic/ Endoplasmic Reticulum Ca^{2+} Transporting 1	8.17E−17	–	–	–	–	–	–	–
86	rs749767	16	31113086	downstream-500B	BCKDK	Branched Chain Ketoacid Dehydrogenase Kinase	1.21E−09	–	–	–	–	–	–	–
87	rs7186084	16	68782357	intron	CDH1	Cadherin 1	–	–	–	–	–	–	–	1.09E−07
88	rs1847591	16	78908913	intron	WWOX	WW domain containing oxidoreductase	–	–	9.99E−07	–	–	–	–	–
89	rs9904685	17	1352101	intron	YWHAE	Tyrosine 3-monooxygenase/ tryptophan 5-monooxygenase activation protein epsilon	6.27E−07	–	–	–	–	–	–	–
90	rs4646404	17	17516885	intron	PEMT	Phosphatidylethanolamine N-Methyltransferase	–	–	5.30E−11	–	–	–	–	–
91	rs9914988	17	28856086	intron	ERAL1	Era like 12S mitochondrial rRNA chaperone 1	–	–	–	–	–	–	–	2.77E−11
92	rs242559	17	45948522	intron	MAPT	Microtubule Associated Protein Tau	–	–	–	8.29E−07	–	–	–	–
93	rs1319247	17	63106279	intron	TANC2	Etratricopeptide repeat, ankyrin repeat and coiled-coil containing 2	–	–	6.02E−08	–	–	–	–	–
94	rs12940622	17	80641771	intron	RPTOR	Regulatory Associated Protein Of MTOR Complex 1	2.49E−09	–	–	–	–	–	–	–
95	rs1044661	17	82943144	intron	B3GNTL1/ TBCD	UDP-GlcNAc:BetaGal Beta-1,3-N-Acetylglucosaminyltransferase Like 1/Tubulin Folding CofactorD			–	–	–	–	–	1.74E−46
96	rs1788785	18	23562376	intron	NPC1	NPC Intracellular Cholesterol Transporter 1	1.98E−08	–	–	–	–	–	–	–

(*Continued on next page*)

Table 4. ***Continued***

No	SNV	Chrom	Position	Role	Gene	Full Gene Name (Cytogenetic Position)	BMI p	OBESITY p	WHR p	GLUC p	INS p	HOMA-B p	HOMAIR p	HbA1c p
97	rs17066842	18	60373391	upstream-2KB	MC4R	Melanocortin 4 Receptor	6.40E−14	–	–	–	–	–	–	–
98	rs12454712	18	63178651	intron	BCL2	BCL2, Apoptosis Regulator	–	–	*1.10E−09*	–	*1.39E−07*	–	–	–
99	rs757318	19	18709498	intron	CRTC1	CREB Regulated Transcription Coactivator 1	8.76E−09	–	–	–	–	–	–	–
100	rs2075650	19	44892362	intron	TOMM40	Translocase Of Outer Mitochondrial Membrane 40	1.25E−08	–	–	–	–	–	–	–
101	rs405509	19	44905579	upstream-2 kb	APOE	Apolipoprotein E	2.65E−07	–	–	–	–	–	–	–
102	rs2281361	20	32140338	intron	TM9SF4	ransmembrane 9 superfamily member 4	9.78E−07	–	–	–	–	–	–	–
103	rs878639	20	35306660	intron	UQCC1	Ubiquinol-Cytochrome C Reductase Complex Assembly Factor 1	–	–	1.50E−11	–	–	–	–	–
104	rs2076574	20	41092733	intron	TOP1	DNA topoisomerase I	–	–	4.11E−07	–	–	–	–	–
105	rs5750373	22	37028990	intron	MPST	Mercaptopyruvate Sulfurtransferase	–	–	–	–	–	–	–	2.17E−07
106	rs2284099	22	43155830	intron	TSPO	Translocator Protein	–	–	6.65E−07	–	–	–	–	–
107	rs1050828	23	154536002	missense	G6PD	Glucose-6-phosphate dehydrogenase	–	–	–	–	–	–	–	8.23E−135
108	rs1448032	23	155052530	intron	FUNDC2	FUN14 domain containing 2	–	–	–	–	–	–	–	5.92E−17
109	rs5940514	23	155559972	intron	TMLHE	Trimethyllysine hydroxylase, epsilon	–	–	–	–	–	–	–	2.22E−19

p values in italics annotate SNVs that associate as significant with more than one trait.

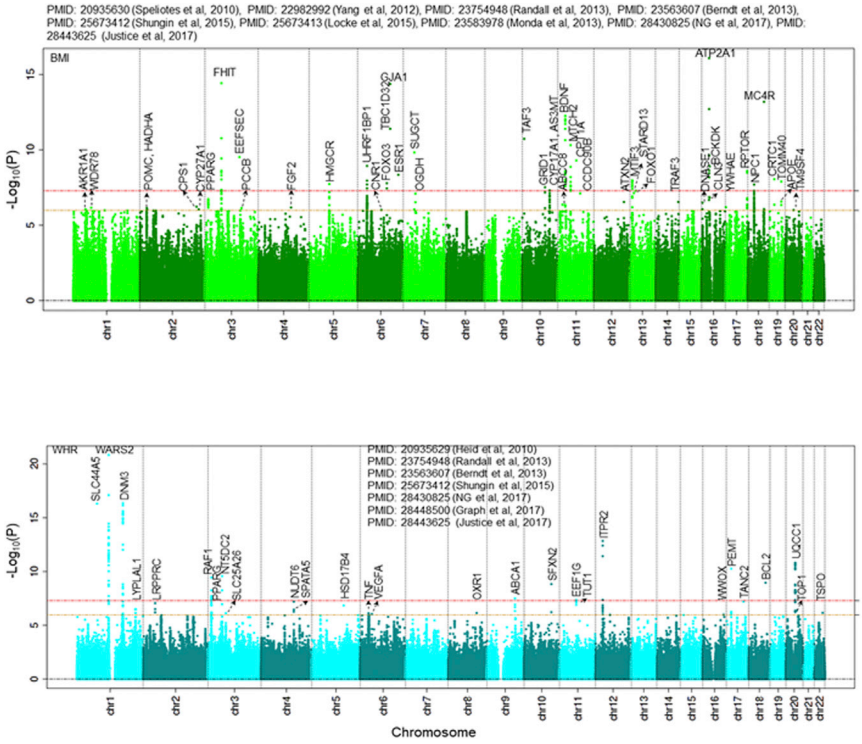


Figure 2. BMI and WHR Association Results with MT-nDNA Candidate Genes

findings demonstrate the importance of several MT-nDNA candidates to cardiometabolic outcomes.

eQTLs of MT-nDNA Candidate Variants

Several variants in or near MT-nDNA candidate genes were identified as expression QTL (eSNV) (Table S8). Based on RegulomeDB,[78] three variants were the best in eSNVs features' ranking. The first was rs242559, intronic to *MAPT* (MIM: 157140). Mutations in *MAPT* associate with lower mitochondrial nicotinamide adenine dinucleotide (NADH) levels,[79] partially suppress complex I-driven respiration, and lower overall ATP production by oxidative phosphorylation, with cells relying on glycolysis to maintain ATP levels. The second, rs9897919, is a 3′ UTR variant for *TBCD* (MIM: 604649), tubulin folding cofactor D, and *B3GNTL1* (17q25.3 [MIM: 615337]), a putative glycosyltransferase. The third, rs1788821, is intronic to *NPC1* (MIM: 607623), which is an intracellular cholesterol transporter with a role in the egress of cholesterol from the endosomal/lysosomal compartments (Table S9).

The findings of RegulomeDB were reinforced by HaploReg (v.4.1).[80] The MT-nDNA variants showed an enrichment in transcription regulation features. For instance, rs242559 of *MAPT* is localized within the promoter histone marks in skeletal muscle, at enhancer histone marks of 16 tissues, and at DNase marks of 4 tissues. In addition, the rs242559 polymorphism alters the protein binding site of GATA2, a transcription factor protein that binds in the promoter regions of target genes (Table S9).

To determine the eSNVs' gene targets in specific tissues, we used GTEx (v.7.0)[81,82] with a summary in Table S8 and detailed in Table S10. The 42 unique eSNVs were assigned to 29 genes targeting regulation of 50 genes, distributed among 13 tissues (adipose, tibial artery, thyroid, skin, blood, brain, skeletal muscle, esophageal muscularis, fibroblast, liver, pancreas, testis, and tibial nerve). There were 28 unique lead and 18 unique secondary (LD r^2 was < 0.02 to the target gene's best eSNV) eSNVs identified (see Material and Methods). For example, rs2510344 of *NPC1* regulates *C18orf8* in skin and rs11663558 of *NPC1* regulates its own *NPC1* gene expressed highest in subcutaneous adipose tissue (GTEx data).

trans-eQTLs of MT-nDNA Candidate Variants

We combined the GWAS p value for the MT-nDNA SNVs with additional evidence from *trans*-gene expression regulation for a specific variant using GWAS3D[83] (Figures S12.1–S12.3). The GWAS3D software selected 16 cell types, which included chromosomal looping data (5C or ChIA-PET or Hi-C) and important transcriptional marker data (H3K4me1, H3K27ac, DHSs, EP300, and CTCF).[83] Among several *trans*-eQTLs, for example, rs2881156 ($p = 8.1E{-}07$) of *SLC25A48* (5q31.1 [MIM: 616150], Figure S12.1) associated with obesity and *trans*-regulated expression of three genes: *SAR1B* (5q31.1 [MIM: 607690]) involved in protein transport from the endoplasmic reticulum to the Golgi (mutations in this gene are a cause of chylomicron retention disease [MIM: 246700]);[84] *TRPC7* (5q31.1), a regulator of intracellular calcium levels;[85] and *REEP2* (5q31.2 [MIM: 609347]), which enhances the function of sweet taste receptors[86] and is about 80 times higher expressed in brain than in other tissues (GTEx data).

PPI Network

We analyzed 109 MT-nDNA proteins using the PPI network (see Web Resources)[87,88] to identify 4,132 interacting proteins. We present the 15 top genes (Table S11) with highest PageRank score for PPI,[89–91] including the number of PPI, a short description of gene's function from Gene Entrez of NCBI-db, associated trait(s), and association p value(s) (see also Supplemental Material and Methods, section 9). From the PPI analysis it is evident that a gene/protein hub (which is assumed important because of a relatively large number of interactions) is not necessary, the top-notch for association with a specific trait, as shown in the Discussion section.

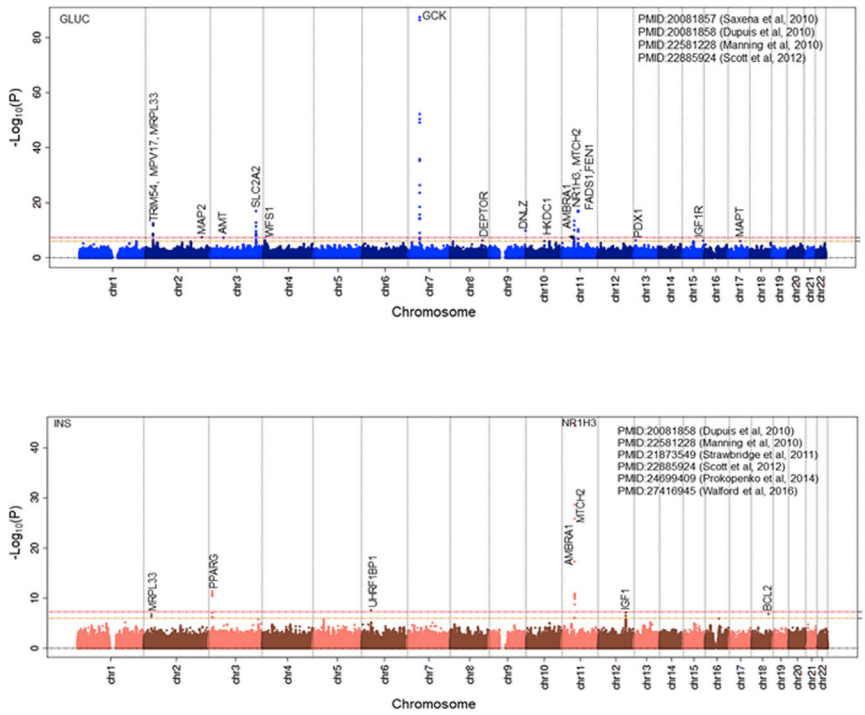


Figure 3. Glucose and Insulin Association Results with MT-nDNA Candidate Genes

Discussion

This large, comprehensive study of human mitochondrial and nuclear mitochondrial genetic variation in relation to seven key risk factors for metabolic disease provides important information toward a better understanding of the causes and mechanisms of these phenotypes. The use of a two-tiered approach—i.e., examining mtDNA variants for up to ~170,202 individuals and nuclear MT-nDNA candidate gene polymorphisms for up to ~339,000 individuals—facilitated the evaluation of the genetic underpinnings of seven metabolic traits reflecting adiposity/obesity and glucose-insulin metabolism and signaling. We identified two mtDNA SNVs (in *MT-ATP6* and the *D-loop*) associated with WHR and fasting glucose, a burden of SNVs in a gene (*MT-ND3*) associated with BMI, and five rare SNVs of mtDNA (in *MT-ATP6*, *MT-ND5*, and *MT-ND6*) associated with BMI (2 of 3), with WHR (2 of 3), and with fasting insulin (1 of 3). Additionally, the MT-nDNA meta-analysis implicated significant associations with 62 of 109 protein coding candidate genes ($p \leq 5E{-}08$), 29 additional passing the above threshold for the metabolic traits considered herein or any other traits querying the GWAS catalog (see Web Resources). Only 16% of MT-nDNA findings passed $p \leq 1E{-}06$ but did not reach $p \leq 5E{-}08$, and thus may be considered as new MT-nDNA contributions in association with seven metabolic traits (Tables 4 and S12). Enrichment analysis showed that 13 of MT-nDNA genes have also been identified as candidate genes for T2D risk and 18 for CVD risk. The eQTL analysis of MT-nDNA candidates revealed that approximately 27% of associations with seven metabolic traits also act as eSNVs in regulating expression of other *cis*-genes. In addition, *trans*-eQTL analysis yielded potential support for maternal and X chromosome epigenetic gene expression regulation. The PPI network analysis identified top PageRank-ed interacting genes from our significant MT-nDNA associations, but they did not necessarily represent the highest trait-specific associations. Overall, these findings indicate the important role of both the nuclear genome and mtDNA-related variants in energy production and, in turn, in the polygenic architecture of metabolic traits. Below we summarize and discuss our most salient study findings.

mtDNA and Metabolic Disease Associations

To date, nearly 290 genetic causes for rare disorders of mitochondrial energy generation have been identified.[92] These rare mitochondrial genetic diseases commonly result in multiple clinical phenotypes with varying penetrance, likely due to differences in heteroplasmy.[93] In our study, *MT-ATP6* variants MT-8706 and MT-8896 (rs202120082), one with MAF > 1% and the other one rare, respectively, were both associated with WHR. *MT-ATP6* is part of the proton-transporting portion of ATP synthase, contributing to its rotational mechanism.[94,95] rs202120082 (MT-8896) is a missense mutation (p.Ala124Thr) and conversion of the hydrophobic alanine to an uncharged, polar threonine at this site may have implications for ATP6's ability to effectively pump protons and thus decrease production of ATP energy units.

MT-16320 (rs62581338; MAF > 1%), in the *D-loop*, which contains a few mtDNA replication origins and is possibly involved in accelerating mtDNA synthesis to satisfy developmental, physiological, or aging-related demands,[96] displayed a significant association with glucose. MT-16320 is less than 1 kb from a major mtDNA transcription initiation site (IT) that overlaps with promoters (in the D-loop in H-strand, IT_{H1} at 561 bps and IT_{H2} at 630, and in L-strand at 407 bps (IT_{L1}))[97] and is 500 bps downstream of mitochondrial-encoded cytochrome b (*MT-CYB* [MIM: 516020], plus strand) and 2 kb upstream of *MT-ND6* (minus strand).

Of the five rare SNVs, MT-14272 (rs2853814, missense), MT-14353, and MT-14584 of *MT-ND6* were associated with adiposity-related traits. *MT-ND6* is a core subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase (complex I) and is part of the minimal assembly required for catalysis. Complex I functions in the transfer of electrons from NADH into the respiratory chain, while

coupling the flow of electrons to the pumping of protons. We also demonstrated an association between insulin and rare variant MT-14124 in the closely related *MT-ND5* gene. *MT-ND5* is also a core subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase (complex I). Future studies need to validate these rare variant mtDNA findings.

It is also important to note that we observed several rare gene-based suggestive associations between our metabolic traits ($p \leq 5E{-}02$, Tables 2 and 3) and non-protein coding genes, primarily tRNAs. Mutations within tRNAs that impact translation would impact all of the mtDNA-encoded peptides and consequently would be expected to have significant effects on oxidative phosphorylation. The *MT-ND3* gene displayed an aggregate of low-frequency variants associated with BMI. *MT-ND3* is a subunit of the respiratory chain complex I that is part of the minimal assembly of core proteins required to catalyze NADH dehydrogenation and electron transfer to ubiquinone (coenzyme Q10). Interestingly, previous studies have demonstrated increased *MT-ND3* gene expression associated with a higher histological severity of hepatic steatosis.[98] Gene-based meta-analysis burden test, which employed inverse variance weighting to combine association results from 82 variants in MT-ND3 gene across cohorts, yielded a significant association between MT-ND3 and BMI ($p = 1E{-}03$, Figure S17). This figure illustrates that rare alleles of 82 SNVs contributing in the MT-ND3 meta-analysis have variable effects.

We used GTEx data for mtDNA genes as assembled by the Human Protein Atlas team (see Web Resources). They showed the highest expression of RNA for the *MT-ATP6*, *MT-ND3*, *MT-ND5*, and *MT-ND6* significant mtDNA genes in the heart and brain, which represent the highest energy-demanding tissues as compared, for example, to adipose tissue, in a ratio of about 3:1 (Figure S13). The importance of MT-nDNA and mtDNA gene polymorphisms in the heart is supported by recent publications that have summarized the relation of MT to vascular function,[99] as therapeutic targets in heart failure,[100] and specific genes and pathways that relate to the heart.[101,102]

MT-nDNA Candidate Genes Associations (Glycemic Traits)

Glucose sensing in β cells is largely controlled by the hexokinase proteins. Three of the genes encoding hexokinases, *HK1* (MIM: 142600), *HKDC1* (MIM: 617221), and *GCK*, were part of our MT-nDNA candidate list and were found to be significantly associated with glucose, HOMA-B, and/or HbA1c (Table 4). Hexokinases catalyze the conversion of glucose into glucose-6-phosphate. The phosphorylation of glucose directly couples extra-mitochondrial glycolysis to intra-mitochondrial oxidative phosphorylation. *GCK*, for example, with pleiotropic effects, produces an enzyme in the pancreas which plays a role in glucose-stimulated insulin secretion and affects glucose uptake and conversion to glycogen in the liver. The *GCK* GWAS-identified variant rs2908289 (Table 4) has been previously associated with glucose, HOMA-B, HbA1c,[54,60] and other variants in LD with it, more recently associated with BMI (rs4607517, $r^2 = 0.65$, $p = 8E{-}56$),[55] with T2D (rs1799884, $r^2 = 0.81$, $p = 5E{-}18$),[103] and with metabolic syndrome (rs3757840, $r^2 = 0.18$, $p = 4E{-}13$).[104] Mitochondrial metabolism, which drives the respiratory chain to produce ATP via oxidative phosphorylation, also contributes to glucose sensing, since disruption of mitochondrial oxidative metabolism blocks glucose-stimulated insulin secretion.[40,105–107] Our analysis shows that 13 additional MT-nDNA candidate genes are annotated with SNVs that associate with glucose/insulin metabolism (Figure S11). Diabetes mellitus has been associated with maternally inherited mutations in *MT-TK* (MIM: 590060), *MT-TS2* (MIM: 590085), and *MT-TE* (MIM: 590025), where the molecular mechanisms involve impaired translation of mtDNA-encoded proteins.[3] In addition, MT-nDNA candidate genes *POLG* (15q26.1 [MIM: 174763], DNA polymerase gamma); *RRM2B* (8q22.3 [MIM: 604712], a ribonucleotide reductase); *OPA1* (3q29 [MIM: 605290], a nuclear-encoded mitochondrial protein with similarity to dynamin-related GTPases); and *MPV17* (2p23.3 [MIM: 137960], a mitochondrial inner membrane protein) have been associated with diabetes mellitus, largely due to an impairment in mtDNA maintenance.[108] In our study, rs1050828, a missense mutation of *G6PD* (Xq28 [MIM: 305900], 154536002 bps; $p = 8.2E{-}135$) was associated with HbA1c. *G6PD* catalyzes the rate-limiting step of oxidative pentose-phosphate pathways, a route for dissimilation of carbohydrates besides glycolysis. It contributes to the production of NADPH and pentose phosphatases for fatty acid and nucleic acid synthesis (Table S6 and Figure S14). *G6PD* plays an essential role in maintaining health by protecting against oxidative damage.[109,110] The same rs1050828 variant of G6PD has previously been reported as associated with HbA1c lowering in African Americans but not in other populations, leading to misdiagnosis/under-diagnosis of T2D.[61] In the last century, a number of papers reported a maternal excess transmission of T2D,[111–113] suggesting the possibility of an epigenetic transmission of diabetes mediated by the mother. To date, two genes, *KLF14* (7q32.2 [MIM: 609393])[114] and *KCNQ1* (11p15.5-p15.4 [MIM: 607542]),[115,116] have been described as potential candidates that show parent-of-origin effects in association with T2D. However, based on our MT-nDNA list, these two genes are not MT-nDNA candidates. In our study, GWAS3D analysis identified *G6PD* as *trans*-regulated with the expression of *MECP2* (Xq28 [MIM: 300005]) (Figure S12.3).[117] *MECP2* binds to methylated DNA and mediates transcriptional repression through interaction with histone deacetylase and the corepressor *SIN3A* (15q24.2 [MIM: 607776]) (Figure S14). Lai et al.[118] reported meta-results that T2D-affected case subjects with *G6PD* deficiency had two times higher odds of developing diabetes than unaffected control subjects.[119] Furthermore,

G6PD protein has a weak score of 1 out of 5 to be found in MT. G6PD is mainly localized in the cytosol and also to the microtubule-organizing center and vesicles (Human Protein Atlas). G6PD protein has 69 interactions and was ranked 65th by the PageRank algorithm for the importance of PPI. These results taken together may suggest that *G6PD* contributes to the maternal epigenetics of T2D, through X chromosome inheritance and less through MT. Other candidates of MT-nDNA genes that may contribute to mitochondrial epigenetics[120] are *DNMT1* (19p13.2 [MIM: 126375], DNA methyltransferase 1), *POLRMT* (19p13.3 [MIM: 601778], RNA polymerase mitochondrial), *TFB1M* (6q25.3 [MIM: 607033], transcription factor B1, mitochondrial), *TFB2M* (1q44 [MIM: 607055], transcription factor B2, mitochondrial), and *TFAM* (10q21.1 [MIM: 600438], transcription factor A, mitochondrial), but showed no significant associations with existing data for the traits studied (for BMI: rs6926853, *TFB1M*, p = 1.3E−03;[45] WHR: rs10465617, *TFB2M*, p = 3.1E−03;[43] glucose: rs4804124, *DNMT1*, p = 7.3E−05;[54] insulin: rs7253062, *DNMT1*, p = 4.4E−03;[55] HOMA-B: rs12462004, *DNMT1*, p = 4.4E−03;[54] HOMA-IR: rs892189, *DNMT1*, p = 7E−04;[54] HbA1c: rs11006132, *TFAM*, p = 2.4E−03[61]).

MT-nDNA Candidate Genes Associations (Adipose Traits)

We found that a total of 26 of the 109 MT-nDNA candidate genes showed significant associations with WHR (Tables 4 and S6 and Figure S11). For example, *WARS2* (1p12 [MIM: 604733]) encodes the MT tryptophanyl-tRNA synthase. We observed an intronic rs6428792 of *WARS2* in association with WHR (p = 7.95E−18).[50] Another SNV, rs10923724 in LD with rs6428792, r^2 = 0.29, WHR p = 9E−25)[52] upstream of *WARS2* but closer to downstream of *TBX15* (1p12 [MIM: 604127]), modifies *WARS2* expression in skeletal muscle (p = 1.1E-36) and in adipose-subcutaneous tissues (p = 1E−s29) (GTEx data). In our PPI analysis of 109 MT-nDNA genes and 4,132 interactants, *WARS2* showed two interactions and was ranked 2,028th based on the PageRank algorithm. Taken together, these findings (our study, GTEx, and GWAS) suggest that associations with WHR might be mediated also by differential expression of *WARS2*.[121]

There were 48 MT-nDNA candidates associating with BMI. For example, *NPC1* (18q11.2, Tables 4 and S6) has been associated with early childhood onset and adult morbid obesity.[122] *NPC1* (a cholesterol transporter) is highly expressed in human white adipose tissue adipocytes with increased levels in obese individuals.[123] Our lead SNV rs1788785 is an eSNV for *NPC1* and is not in LD with the best *NPC1* eSNV, supporting *NPC1* as a candidate gene for obesity. Indeed, studies in mice have shown that a non-functioning *NPC1* resulted in late-onset weight loss and less food intake.[124] In humans, *NPC1* is part of the cholesterol metabolism pathway (see Web Resources). Recently, *NPC1* mutant cells were reported to have fragmented mitochondrial networks, increased respiration, alterations in the composition of the respiratory chain complex, and a substantial reduction in the cellular ATP level. Thus, a primary lysosomal defect in *NPC1* mutant fibroblasts is accompanied by deregulation of the organization and function of the mitochondrial network.[125] *NPC1* was ranked 241st and had 13 interactions in the PPI network.

Other notable obesity-related variants/genes among MT-nDNA candidates include *POMC* (2p23.3, adrenocorticotropic peptide/hormone, p = 7.2E−07), which binds to melanocortin 2 receptor (*MC2R*, 18p11.21 [MIM: 607397]), stimulating release of cortisol, a steroid stress-response hormone. *MC4R* (18q21.32, p = 6.4E−14) was ranked the 3rd in BMI-effects compared to *FTO* (16q12.2 [MIM: 610966]), the top ranked for BMI effects in Speliotes et al.,[41] Locke et al.,[46] and Winkler et al.[48] *MC4R* is another member of the melanocortin receptor family, which has a central role in energy homeostasis and somatic growth. *MC4R* is ranked the 82nd, with 12 PPI, when analyzed for the importance of PPI using the PageRank algorithm. *CNR1* (6q15 [MIM: 114610], cannabinoid receptor, p = 9.6E−07) influences mitochondrial respiration. *CRTC1* (19p13.11 [MIM: 607536], CREB regulated transcription coactivator 1, p = 8.8E−09) is a potent coactivator of *PGC1a* (4p15.2, officially known as PPARGC1A [MIM: 604517]), a transcriptional coactivator that regulates the genes involved in energy metabolism, and inducer of mitochondrial biogenesis.

MTCH2 (11p11.2, mitochondrial carrier homolog 2, rs7118178, BMI p = 5.1E−08) was 9th in BMI effects compared to *FTO* in Speliotes et al.[41] and 16th in Locke et al.[46] MTCH2 is located in the inner membrane of MT, highly expressed in white adipose tissue and adipocytes, and is thought to play a regulatory role in adipocyte differentiation and biology.[126] The *MTCH2* GWAS-variant rs7118178 (Table 4) has been associated with BMI,[46] glucose,[36,55] and pro-insulin[58] (Figure 4, T1-4). *MTCH2* has 87 protein interactions and ranked 14th when analyzed for the importance of PPI using PageRank algorithm. *MTCH2* deficiency in mouse muscle has been shown to be beneficial, protecting mice from the obesogenic effect of high-fat diets, most likely the result of an increase in mitochondrial metabolism. *MTCH2* is proposed as a repressor of muscle mitochondrial metabolism and size.[127] Using both GWAS and eQTL information, three *MTCH2* SNVs co-localize within a 15 kbp DNA region. Based on GTEx data, rs4752856 in *MTCH2* (Figure 4, T6) regulates gene expression of *C1QTNF4* (11p11.2 [MIM: 614911]) and *PSMC3* (11p11.2 [MIM: 186852]) in subcutaneous adipose tissue. rs3817335, which is in high LD (r^2 = 0.903) with rs4752856, regulates *SLC39A13* (11p11.2 [MIM: 608735]) in tibial artery tissue, while rs7118178 (in less LD to two other SNVs, r^2 = 0.083) regulates *CELF1* (11p11.2 [MIM: 601074]) in tibial nerve tissue, and with results similar to the best eSNVs for the mentioned genes and in high LD with their eSNVs (Figure 4, T5). *C1QTNF4* is reported as a potential cytokine

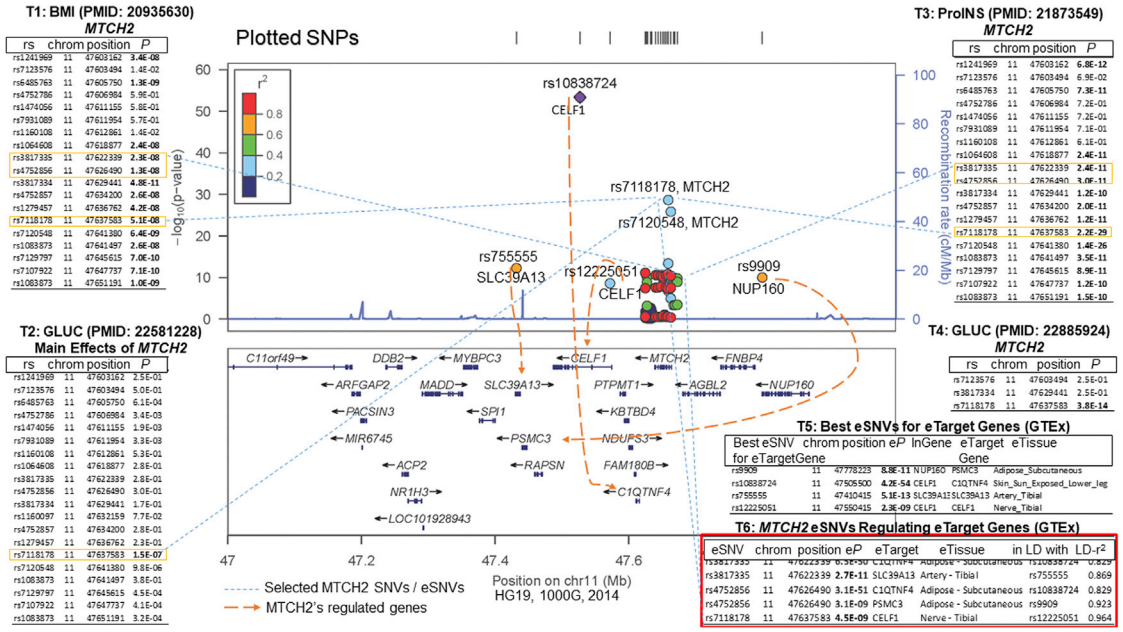


Figure 4. Selected SNVs of *MTCH2* Significantly Associating with BMI, Glucose, Insulin, and a Few of Them Regulating the Expression (eSNVs) of at Least Four Genes: *C1QTNF4, SLC39A13, PSMC3,* and *CELF1*

First, we used LocusZoom[142] for plotting regional information for *MTCH2* selected SNVs, which have been reported by large GWAS publications (for BMI[41] [T1], for glucose[55] [T2], for proinsulin[58] [T3], and another large analysis for glucose[56] [T4]), for strength and extent of the association signal relative to genomic position, local linkage disequilibrium, and recombination patterns and the positions of genes in the region. Second, we used GTEx[81] data to identify the SNVs of *MTCH2* that are eSNVs by influencing the expression of other *cis*-target-genes (*C1QTNF4*, *PSMC3*, *SLC39A13*, *CELF1*) (T6). Third, we verified whether these eSNVs of *MTCH2* (connected with blue dashed lines) are true expression regulators, by identifying the best eSNVs of the four target genes (T5, regulation effects shown with orange-dashed lines) and evaluating the LD r^2 between the *MTCH2* eSNV, e.g., rs4752856 (T6) and the best eSNV rs10838724 (T5) for the corresponding target gene (*C1QTNF4*), which was found to be 0.829 in adipose-subcutaneous tissue (T6). Fourth, the *MTCH2* SNVs serving as eSNVs for the four *cis*-target-genes, shown in T6, colocalize in about 15K bps region. Thus, *MTCH2* associates with BMI, glucose, and proinsulin, and among others influence the gene expression of other *cis*-genes (*C1QTNF4*, *PSMC3*, *SLC39A13*, *CELF1*), who in itself contribute also to individuals' obesity risk (see Discussion).

that can induce the activation of both *NFKB1* (MIM: 164011) and *IL6/STAT3* (MIM: 147620/102582) signaling pathways[128] and acts in the hypothalamus to modulate food intake and peripheral energy expenditure.[129] *PSMC3* is involved in ATP-dependent degradation of misfolded and damaged proteins and in removal of no longer required proteins. This protein is also involved in cell cycle progression, apoptosis, and DNA damage repair. A knockdown of *PSMC3* in human immortalized fibroblasts increased cell proliferation.[130] *CELF1* has been associated with Alzheimer disease and obesity.[131] Consequently, *MTCH2* SNVs regulate gene expression of other *cis*-genes related to obesity, food intake, and/or cell number. Thus, *MTCH2* effects on BMI may be more complex than previously accounted for.

Other Functions of MT-nDNA Candidate Genes

Mitochondrial inheritance has been previously associated with obesity, metabolic syndrome, insulin resistance, type 2 diabetes, and cardiovascular disease.[3,40,132–135] The complex genetics of MT affects processes of glucose, lipid, and amino acid catabolism, with ATP as the final energy product. By-products of these processes play an important role in signaling (e.g., ROS) and are used for epigenetic modifications (e.g., acetyl CoA). In our study, we would expect a preponderance of glycemic and lipid genes and proteins. Out of 109 MT-nDNA candidate genes, 21 of the significantly associated genes were functionally related with cholesterol (Table S6 and Figure S11), 16 with glucose and insulin, and 5 with adipose/obesity. The 109 MT-nDNA candidate genes have other functions too. For example, rs1044661 is located at 17q25.3, overlapping two genes that run in opposite strands: *TBCD* (involved in tubulin folding as cofactor D) and *B3GNTL1* (a putative glucosyltransferase). This SNV was associated with HbA1c ($p = 1.74E{-}46$).[61] Recently, Francis et al.[136] hypothesized that the interaction between *TBCD* (17q25.3) and *ARL2* (11q13.1 [MIM: 601175]) suggest that *ARL2* serves as regulator of both mitochondrial fusion and microtubule dynamics. In 109 MT-nDNA candidates, *WDR78* (1p31.3), *TRIM54* (2p23.3 [MIM: 606474], regulating titin kinase [controlling elasticity in muscle] and microtubule-dependent signaling pathways in striated muscles), *DNM3* (1q24.3 [MIM: 611445], involved in vesicular transport and with biased expression in brain), *MAP2* (2q34 [MIM: 157130], involved in microtubule assembly and neurogenesis), *GJA1* (6q22.31 [MIM: 121014], involved in the

contraction of heart) and *MAPT* (17q21.31, expressed in nervous system) produce microtubule-associated proteins in interaction with MT (Table S6 and Figure S11). It should be noted that rs1050828, a missense mutation of *G6PD* residing on X chromosome, may impact HbA1c through non-glycemic factors rather than glucose metabolism.[61] Additional functions of MT-nDNA candidate genes are described in the Supplemental Material and Methods, section 7.

Strengths and Limitations

Our study had many strengths. We used large sample sizes for studying mtDNA and MT-nDNA. At the mtDNA level we validated SNVs used per cohort and imputed based on Cosmopolitan MT-1000 Genomes. At the MT-nDNA level we used the strength of many consortia contributions for associations with seven traits. As a consequence, we performed comprehensive mtDNA and MT-nDNA analyses. The seven mtDNA variants significant associations showed no statistical significant deviations from the homogeneity/similarity of beta contributions before meta-pooling across studies as reported with Het-p in Table 1, as well as graphically represented in the Forest plot (Figure S16). For supporting MT-future studies, we have provided mtDNA updated Tables S15–S28 for arrays used in this study including MitoMap annotation. Also, we include Table S32 of predictions from the bioinformatics platforms that have been shown to have the highest sensitivity and specificity for the mtDNA variants.[137] Our study also had some limitations. Rare mtDNA mutations are usually heteroplasmic point mutations and/or mtDNA lesions that typically result in a primary mitochondrial disease, which can manifest as a broad range of clinical outcomes. In this study we have interrogated the effect of inherited mtDNA variants, much commoner in the population than typical mtDNA mutations. Therefore, while there is a great deal of literature linking inherited mtDNA variants to disease, many of these studies have limited statistical power[138] and several have never been independently replicated. So in many respects there is not a great deal of robust literature to call upon. The harmonization of mtDNA in each study was dependent on the number of mtDNA variants and their quality. Some variants were dropped before performing any imputation, because low-quality variants matched not only to the mtDNA genome, but also to nuclear genome (see mtDNA Variant Harmonization in Material and Methods). A number of quality-control filters were applied to mtDNA, which reduced individual sample size per marker or the number of useful markers. We did not address the association of significant markers by gender, because in our study sample sizes per mtDNA marker were variable and often small.

Perspective and Conclusion

We focused on associations of MT with adipose and glycemic traits, yet we acknowledge that there is still much to be learned from other traits, for example for triglycerides, high-density lipoprotein cholesterol, and C-reactive protein.[46,139–141] Notably, we identified 21 MT genes of lipid metabolism, although this was not the direct focus of our study. For example, *TMLHE* (Xq28 [MIM: 300777], trimethyllysine dioxygenase associated with HbA1c, $p = 2.2E{-}19$) is the first enzyme in the carnitine biosynthesis pathway. Carnitine play an essential role in the transport of activated fatty acids across the inner mitochondrial membrane and this gene is ubiquitous for its expression in heart. Hence, MT associations with other traits at the consortia-level remain to be explored. In conclusion, we identified common, rare SNVs, and a gene burden of genetic mtDNA variants, as well as 109 MT-nDNA genes associated with metabolic traits. Of the 109 MT-nDNA candidate genes, a subset pointed to associations with adipose, glucose metabolism, T2D, and CVD, and a subset of SNVs was inferred to contribute for differential regulatory genes expression. We documented in this study that the MT-candidates (SNVs/genes) have special contributions for functioning and energy homeostasis in adipose tissues and to glucose metabolism and insulin signaling.

Accession Numbers

Summary statistics for the mtDNA meta-analyses are available in the dbGaP repository for the CHARGE Consortium, phs000930.v8.p1.

Supplemental Data

Supplemental Data include 17 figures, 32 tables, Supplemental Material and Methods, Supplemental Acknowledgments, and Supplemental Study Descriptions and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.12.001.

Acknowledgments

A.E. Justice is supported by a K99/R00 from the National Institutes of Health, National Heart, Lung, and Blood Institute (NHLBI) (5K99HL130580-02 and 5R00HL130580-04). G. Hudson has received funding support by Wellcome Trust Centre for Mitochondrial Research, Grant Code G906919. J.I.R., X.G., Y.-D.I.C., K.D.T., Y.H., K.S., and J.Y. have received NIH support for research related to mitochondria. A.T.K., C. Liu, J.L.F., M. Graff, C.T.H., C.G., L.R.Y., M.F.F., D.E.A., D.I.C., K.Y., W.D.H., S.J.L., L.W., T.J.B., L.D.L.-G., T.I.A.S., G. Hudson, D.L., A.N.M., D.M.B., M.A.P., J.B.M., J.I.R., and K.E.N. were the writing group for this paper. M.A.P. and K.E.N. are from the CHARGE ADIPOSITY working group; J.I.R. and J.B.M. are from the CHARGE DIABETES working group.

Declaration of Interests

A.Y.C. is currently an employee of Merck & Co. C.T.-P. has received grants from Bayer pertaining to antithrombotic treatment of atrial fibrillation and cardiac valve replacements. D.O.M.-K. works as a part-time clinical research consultant for Metabolon, Inc. The remaining authors declare no competing financial interests.

Received: July 1, 2018
Accepted: December 6, 2018
Published: December 27, 2018

Web Resources

1000 Genomes data (released April 2016), http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/
dbGaP, https://www.ncbi.nlm.nih.gov/gap
EasyQC, http://www.genepi-regensburg.de/easyqc/
FCGENE, https://sourceforge.net/projects/fcgene/
GenBank (NCBI Reference Sequence: NC_012920.1), https://www.ncbi.nlm.nih.gov/nuccore/251831106
GIANT Consortium repositories, https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files
GTEx (Gene Expression Portal), https://gtexportal.org
GWAS Catalog, http://www.ebi.ac.uk/gwas/
Human Mitochondrion Sequence, https://www.mitomap.org/MITOMAP/HumanMitoSeq
IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html
KEGG, https://www.genome.jp/dbget-bin/www_bget?hsa04979
MAGIC consortium archives, https://www.magicinvestigators.org/downloads/
OMIM, http://www.omim.org/
Protein-Protein Interactions, ftp://ftp.ncbi.nih.gov/gene/GeneRIF/
seqMeta, https://github.com/DavisBrian/seqMeta
seqMeta release notes, https://cran.r-project.org/web/packages/seqMeta/seqMeta.pdf
SHAPEIT2, https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html#home
The Human Protein Atlas, http://www.proteinatlas.org/

References

1. Wang, Z., Ying, Z., Bosy-Westphal, A., Zhang, J., Schautz, B., Later, W., Heymsfield, S.B., and Müller, M.J. (2010). Specific metabolic rates of major organs and tissues across adulthood: evaluation by mechanistic model of resting energy expenditure. Am. J. Clin. Nutr. *92*, 1369–1377.
2. Ernster, L., and Schatz, G. (1981). Mitochondria: a historical review. J. Cell Biol. *91*, 227s–255s.
3. Chow, J., Rahman, J., Achermann, J.C., Dattani, M.T., and Rahman, S. (2017). Mitochondrial disease and endocrine dysfunction. Nat. Rev. Endocrinol. *13*, 92–104.
4. Prasai, K. (2017). Regulation of mitochondrial structure and function by protein import: A current review. Pathophysiology *24*, 107–122.
5. Lang, B.F., Gray, M.W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. Annu. Rev. Genet. *33*, 351–397.
6. Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S.A., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K., et al. (2008). A mitochondrial protein compendium elucidates complex I disease biology. Cell *134*, 112–123.
7. Smith, A.C., and Robinson, A.J. (2016). MitoMiner v3.1, an update on the mitochondrial proteomics database. Nucleic Acids Res. *44* (D1), D1258–D1261.
8. Febbo, P.G., Mulligan, M.G., Slonina, D.A., Stegmaier, K., Di Vizio, D., Martinez, P.R., Loda, M., and Taylor, S.C. (2007). Literature Lab: a method of automated literature interrogation to infer biology from microarray analysis. BMC Genomics *8*, 461.
9. Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B., and Keightley, P.D. (2008). Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. PLoS Biol. *6*, e204.
10. Neiman, M., and Taylor, D.R. (2009). The causes of mutation accumulation in mitochondrial genomes. Proc. Biol. Sci. *276*, 1201–1209.
11. Craven, L., Alston, C.L., Taylor, R.W., and Turnbull, D.M. (2017). Recent advances in mitochondrial disease. Annu. Rev. Genomics Hum. Genet. *18*, 257–275.
12. Schaefer, A.M., Walker, M., Turnbull, D.M., and Taylor, R.W. (2013). Endocrine disorders in mitochondrial disease. Mol. Cell. Endocrinol. *379*, 2–11.
13. Malik, A.N., and Czajka, A. (2013). Is mitochondrial DNA content a potential biomarker of mitochondrial dysfunction? Mitochondrion *13*, 481–492.
14. Ashar, F.N., Moes, A., Moore, A.Z., Grove, M.L., Chaves, P.H.M., Coresh, J., Newman, A.B., Matteini, A.M., Bandeen-Roche, K., Boerwinkle, E., et al. (2015). Association of mitochondrial DNA levels with frailty and all-cause mortality. J. Mol. Med. (Berl.) *93*, 177–186.
15. Mengel-From, J., Thinggaard, M., Dalgård, C., Kyvik, K.O., Christensen, K., and Christiansen, L. (2014). Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. Hum. Genet. *133*, 1149–1159.
16. Ashar, F.N., Zhang, Y., Longchamps, R.J., Lane, J., Moes, A., Grove, M.L., Mychaleckyj, J.C., Taylor, K.D., Coresh, J., Rotter, J.I., et al. (2017). Association of mitochondrial DNA copy number with cardiovascular disease. JAMA Cardiol. *2*, 1247–1255.
17. Zhang, Y., Guallar, E., Ashar, F.N., Longchamps, R.J., Castellani, C.A., Lane, J., Grove, M.L., Coresh, J., Sotoodehnia, N., Ilkhanoff, L., et al. (2017). Association between mitochondrial DNA copy number and sudden cardiac death: findings from the Atherosclerosis Risk in Communities study (ARIC). Eur. Heart J. *38*, 3443–3448.
18. Tin, A., Grams, M.E., Ashar, F.N., Lane, J.A., Rosenberg, A.Z., Grove, M.L., Boerwinkle, E., Selvin, E., Coresh, J., Pankratz, N., and Arking, D.E. (2016). Association between mitochondrial DNA copy number in peripheral blood and incident CKD in the Atherosclerosis Risk in Communities Study. J. Am. Soc. Nephrol. *27*, 2467–2473.
19. Heinonen, S., Buzkova, J., Muniandy, M., Kaksonen, R., Ollikainen, M., Ismail, K., Hakkarainen, A., Lundbom, J., Lundbom, N., Vuolteenaho, K., et al. (2015). Impaired mitochondrial biogenesis in adipose tissue in acquired obesity. Diabetes *64*, 3135–3145.
20. De Pauw, A., Tejerina, S., Raes, M., Keijer, J., and Arnould, T. (2009). Mitochondrial (dys)function in adipocyte (de)differentiation and systemic metabolic alterations. Am. J. Pathol. *175*, 927–939.
21. Altshuler-Keylin, S., and Kajimura, S. (2017). Mitochondrial homeostasis in adipose tissue remodeling. Sci. Signal. *10*, 10.
22. Montgomery, M.K., and Turner, N. (2015). Mitochondrial dysfunction and insulin resistance: an update. Endocr. Connect. *4*, R1–R15.
23. Vernochet, C., Damilano, F., Mourier, A., Bezy, O., Mori, M.A., Smyth, G., Rosenzweig, A., Larsson, N.G., and Kahn,

C.R. (2014). Adipose tissue mitochondrial dysfunction triggers a lipodystrophic syndrome with insulin resistance, hepatosteatosis, and cardiovascular complications. FASEB J. *28*, 4408–4419.
24. Wiklund, P., Zhang, X., Pekkala, S., Autio, R., Kong, L., Yang, Y., Keinänen-Kiukaanniemi, S., Alen, M., and Cheng, S. (2016). Insulin resistance is associated with altered amino acid metabolism and adipose tissue dysfunction in normoglycemic women. Sci. Rep. *6*, 24540.
25. Beasley, T.M., Erickson, S., and Allison, D.B. (2009). Rank-based inverse normal transformations are increasingly used, but are they merited? Behav. Genet. *39*, 580–595.
26. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.
27. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
28. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.
29. Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods *10*, 5–6.
30. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
31. Delaneau, O., Marchini, J.; and 1000 Genomes Project Consortium (2014). Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat. Commun. *5*, 3934.
32. Winkler, T.W., Day, F.R., Croteau-Chonka, D.C., Wood, A.R., Locke, A.E., Mägi, R., Ferreira, T., Fall, T., Graff, M., Justice, A.E., et al.; Genetic Investigation of Anthropometric Traits (GIANT) Consortium (2014). Quality control and conduct of genome-wide association meta-analyses. Nat. Protoc. *9*, 1192–1212.
33. Chen, H., Meigs, J.B., and Dupuis, J. (2013). Sequence kernel association test for quantitative traits in family samples. Genet. Epidemiol. *37*, 196–204.
34. Liu, C., Dupuis, J., Larson, M.G., and Levy, D. (2013). Association testing of the mitochondrial genome using pedigree data. Genet. Epidemiol. *37*, 239–247.
35. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
36. Huang, H., Chanda, P., Alonso, A., Bader, J.S., and Arking, D.E. (2011). Gene-based tests of association. PLoS Genet. *7*, e1002177.
37. Wang, L., Lee, S., Gim, J., Qiao, D., Cho, M., Elston, R.C., Silverman, E.K., and Won, S. (2016). Family-based rare variant association analysis: A fast and efficient method of multivariate phenotype association analysis. Genet. Epidemiol. *40*, 502–511.
38. Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. Am. J. Hum. Genet. *95*, 5–23.
39. Calvo, S.E., Clauser, K.R., and Mootha, V.K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. *44* (D1), D1251–D1257.
40. Lu, H., Koshkin, V., Allister, E.M., Gyulkhandanyan, A.V., and Wheeler, M.B. (2010). Molecular and metabolic evidence for mitochondrial defects associated with beta-cell dysfunction in a mouse model of type 2 diabetes. Diabetes *59*, 448–459.
41. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Mägi, R., et al.; MAGIC; and Procardis Consortium (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat. Genet. *42*, 937–948.
42. Yang, J., Loos, R.J., Powell, J.E., Medland, S.E., Speliotes, E.K., Chasman, D.I., Rose, L.M., Thorleifsson, G., Steinthorsdottir, V., Mägi, R., et al. (2012). FTO genotype is associated with phenotypic variability of body mass index. Nature *490*, 267–272.
43. Berndt, S.I., Gustafsson, S., Mägi, R., Ganna, A., Wheeler, E., Feitosa, M.F., Justice, A.E., Monda, K.L., Croteau-Chonka, D.C., Day, F.R., et al. (2013). Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat. Genet. *45*, 501–512.
44. Randall, J.C., Winkler, T.W., Kutalik, Z., Berndt, S.I., Jackson, A.U., Monda, K.L., Kilpeläinen, T.O., Esko, T., Mägi, R., Li, S., et al.; DIAGRAM Consortium; and MAGIC Investigators (2013). Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. PLoS Genet. *9*, e1003500.
45. Monda, K.L., Chen, G.K., Taylor, K.C., Palmer, C., Edwards, T.L., Lange, L.A., Ng, M.C., Adeyemo, A.A., Allison, M.A., Bielak, L.F., et al.; NABEC Consortium; UKBEC Consortium; BioBank Japan Project; and AGEN Consortium (2013). A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. Nat. Genet. *45*, 690–696.
46. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al.; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; and International Endogene Consortium (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature *518*, 197–206.
47. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Mägi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., et al.; ADIPOGen Consortium; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GEFOS Consortium; GENIE Consortium; GLGC; ICBP; International Endogene Consortium; LifeLines Cohort Study; MAGIC Investigators; MuTHER Consortium; PAGE Consortium; and ReproGen Consortium (2015). New genetic loci link adipose and insulin biology to body fat distribution. Nature *518*, 187–196.
48. Winkler, T.W., Justice, A.E., Graff, M., Barata, L., Feitosa, M.F., Chu, S., Czajkowski, J., Esko, T., Fall, T., Kilpeläinen, T.O., et al.; CHARGE Consortium; DIAGRAM Consortium; GLGC Consortium; Global-BPGen Consortium; ICBP Consortium; and MAGIC Consortium (2015). The influence of age and sex on genetic associations with adult body size

and shape: a large-scale genome-wide interaction study. PLoS Genet. *11*, e1005378.

49. Ng, M.C.Y., Graff, M., Lu, Y., Justice, A.E., Mudgal, P., Liu, C.T., Young, K., Yanek, L.R., Feitosa, M.F., Wojczynski, M.K., et al.; Bone Mineral Density in Childhood Study (BMDCS) Group (2017). Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. PLoS Genet. *13*, e1006719.
50. Justice, A.E., Winkler, T.W., Feitosa, M.F., Graff, M., Fisher, V.A., Young, K., Barata, L., Deng, X., Czajkowski, J., Hadley, D., et al. (2017). Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. Nat. Commun. *8*, 14977.
51. Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinthorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Mägi, R., et al.; MAGIC (2010). Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat. Genet. *42*, 949–960.
52. Graff, M., Scott, R.A., Justice, A.E., Young, K.L., Feitosa, M.F., Barata, L., Winkler, T.W., Chu, A.Y., Mahajan, A., Hadley, D., et al.; CHARGE Consortium; EPIC-InterAct Consortium; and PAGE Consortium (2017). Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. PLoS Genet. *13*, e1006528.
53. Saxena, R., Hivert, M.F., Langenberg, C., Tanaka, T., Pankow, J.S., Vollenweider, P., Lyssenko, V., Bouatia-Naji, N., Dupuis, J., Jackson, A.U., et al.; GIANT consortium; and MAGIC investigators (2010). Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat. Genet. *42*, 142–148.
54. Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L., et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; and MAGIC investigators (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat. Genet. *42*, 105–116.
55. Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T., Bielak, L.F., Prokopenko, I., et al.; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; and Multiple Tissue Human Expression Resource (MUTHER) Consortium (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat. Genet. *44*, 659–669.
56. Scott, R.A., Lagou, V., Welch, R.P., Wheeler, E., Montasser, M.E., Luan, J., Mägi, R., Strawbridge, R.J., Rehnberg, E., Gustafsson, S., et al.; DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat. Genet. *44*, 991–1005.
57. Prokopenko, I., Poon, W., Mägi, R., Prasad B, R., Salehi, S.A., Almgren, P., Osmark, P., Bouatia-Naji, N., Wierup, N., Fall, T., et al. (2014). A central role for GRB10 in regulation of islet function in man. PLoS Genet. *10*, e1004235.
58. Strawbridge, R.J., Dupuis, J., Prokopenko, I., Barker, A., Ahlqvist, E., Rybin, D., Petrie, J.R., Travers, M.E., Bouatia-Naji, N., Dimas, A.S., et al.; DIAGRAM Consortium; GIANT Consortium; MuTHER Consortium; CARDIoGRAM Consortium; and C4D Consortium (2011). Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. Diabetes *60*, 2624–2634.
59. Walford, G.A., Gustafsson, S., Rybin, D., Stančáková, A., Chen, H., Liu, C.T., Hong, J., Jensen, R.A., Rice, K., Morris, A.P., et al. (2016). Genome-wide association study of the modified Stumvoll insulin sensitivity index identifies BCL2 and FAM19A2 as novel insulin sensitivity loci. Diabetes *65*, 3200–3211.
60. Soranzo, N., Sanna, S., Wheeler, E., Gieger, C., Radke, D., Dupuis, J., Bouatia-Naji, N., Langenberg, C., Prokopenko, I., Stolerman, E., et al.; WTCCC (2010). Common variants at 10 genomic loci influence hemoglobin A_1(C) levels via glycemic and nonglycemic pathways. Diabetes *59*, 3229–3239.
61. Wheeler, E., Leong, A., Liu, C.T., Hivert, M.F., Strawbridge, R.J., Podmore, C., Li, M., Yao, J., Sim, X., Hong, J., et al.; EPIC-CVD Consortium; EPIC-InterAct Consortium; and Lifelines Cohort Study (2017). Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med. *14*, e1002383.
62. Chanda, P., Huang, H., Arking, D.E., and Bader, J.S. (2013). Fast association tests for genes with FAST. PLoS ONE *8*, e68585.
63. Zhao, W., Rasheed, A., Tikkanen, E., Lee, J.J., Butterworth, A.S., Howson, J.M.M., Assimes, T.L., Chowdhury, R., Orho-Melander, M., Damrauer, S., et al.; CHD Exome+ Consortium; EPIC-CVD Consortium; EPIC-Interact Consortium; and Michigan Biobank (2017). Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. Nat. Genet. *49*, 1450–1457.
64. Spracklen, C.N., Chen, P., Kim, Y.J., Wang, X., Cai, H., Li, S., Long, J., Wu, Y., Wang, Y.X., Takeuchi, F., et al. (2017). Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. Hum. Mol. Genet. *26*, 1770–1784.
65. Nelson, C.P., Goel, A., Butterworth, A.S., Kanoni, S., Webb, T.R., Marouli, E., Zeng, L., Ntalla, I., Lai, F.Y., Hopewell, J.C., et al.; EPIC-CVD Consortium; CARDIoGRAMplusC4D; and UK Biobank CardioMetabolic Consortium CHD working group (2017). Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat. Genet. *49*, 1385–1391.
66. Eppinga, R.N., Hagemeijer, Y., Burgess, S., Hinds, D.A., Stefansson, K., Gudbjartsson, D.F., van Veldhuisen, D.J., Munroe, P.B., Verweij, N., and van der Harst, P. (2016). Identification of genomic loci associated with resting heart rate and shared genetic predictors with all-cause mortality. Nat. Genet. *48*, 1557–1563.
67. Warren, H.R., Evangelou, E., Cabrera, C.P., Gao, H., Ren, M., Mifsud, B., Ntalla, I., Surendran, P., Liu, C., Cook, J.P., et al.; International Consortium of Blood Pressure (ICBP) 1000G Analyses; BIOS Consortium; Lifelines Cohort Study; Understanding Society Scientific group; CHD Exome+ Consortium; ExomeBP Consortium; T2D-GENES Consortium; GoT2DGenes Consortium; Cohorts for Heart and Ageing

Research in Genome Epidemiology (CHARGE) BP Exome Consortium; International Genomics of Blood Pressure (iGEN-BP) Consortium; and UK Biobank CardioMetabolic Consortium BP working group (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat. Genet. *49*, 403–415.

68. Sabater-Lleal, M., Huang, J., Chasman, D., Naitza, S., Dehghan, A., Johnson, A.D., Teumer, A., Reiner, A.P., Folkersen, L., Basu, S., et al.; VTE Consortium; STROKE Consortium; Wellcome Trust Case Control Consortium 2 (WTCCC2); C4D Consortium; and CARDIoGRAM Consortium (2013). Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated Loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. Circulation *128*, 1310–1324.
69. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al.; Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.
70. Mozaffarian, D., Kabagambe, E.K., Johnson, C.O., Lemaitre, R.N., Manichaikul, A., Sun, Q., Foy, M., Wang, L., Wiener, H., Irvin, M.R., et al. (2015). Genetic loci associated with circulating phospholipid trans fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. Am. J. Clin. Nutr. *101*, 398–406.
71. Yeo, A., Li, L., Warren, L., Aponte, J., Fraser, D., King, K., Johansson, K., Barnes, A., MacPhee, C., Davies, R., et al. (2017). Pharmacogenetic meta-analysis of baseline risk factors, pharmacodynamic, efficacy and tolerability endpoints from two large global cardiovascular outcomes trials for darapladib. PLoS ONE *12*, e0182115.
72. Hinds, D.A., Buil, A., Ziemek, D., Martinez-Perez, A., Malik, R., Folkersen, L., Germain, M., Mälarstig, A., Brown, A., Soria, J.M., et al.; METASTROKE Consortium, INVENT Consortium (2016). Genome-wide association analysis of self-reported events in 6135 individuals and 252 827 controls identifies 8 loci associated with thrombosis. Hum. Mol. Genet. *25*, 1867–1874.
73. Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C., Hwang, S.J., et al.; International Consortium for Blood Pressure Genome-Wide Association Studies; CARDI oGRAM consortium; CKDGen Consortium; KidneyGen Consortium; EchoGen consortium; and CHARGE-HF consortium (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature *478*, 103–109.
74. van der Harst, P., van Setten, J., Verweij, N., Vogler, G., Franke, L., Maurano, M.T., Wang, X., Mateo Leach, I., Eijgelsheim, M., Sotoodehnia, N., et al. (2016). 52 genetic loci influencing myocardial mass. J. Am. Coll. Cardiol. *68*, 1435–1448.
75. Pott, J., Burkhardt, R., Beutner, F., Horn, K., Teren, A., Kirsten, H., Holdt, L.M., Schuler, G., Teupser, D., Loeffler, M., et al. (2017). Genome-wide meta-analysis identifies novel loci of plaque burden in carotid artery. Atherosclerosis *259*, 32–40.
76. Yasuno, K., Bilguvar, K., Bijlenga, P., Low, S.K., Krischek, B., Auburger, G., Simon, M., Krex, D., Arlier, Z., Nayak, N., et al. (2010). Genome-wide association study of intracranial aneurysm identifies three new risk loci. Nat. Genet. *42*, 420–425.
77. Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T., et al.; ENGAGE Consortium (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat. Genet. *41*, 47–55.
78. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. *22*, 1790–1797.
79. Esteras, N., Rohrer, J.D., Hardy, J., Wray, S., and Abramov, A.Y. (2017). Mitochondrial hyperpolarization in iPSC-derived neurons from patients of FTDP-17 with 10+16 MAPT mutation leads to oxidative stress and neurodegeneration. Redox Biol. *12*, 410–422.
80. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. *40*, D930–D934.
81. Battle, A., Brown, C.D., Engelhardt, B.E., Montgomery, S.B.; GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; and eQTL manuscript working group (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.
82. Saha, A., Kim, Y., Gewirtz, A.D.H., Jo, B., Gao, C., McDowell, I.C., Engelhardt, B.E., Battle, A.; and GTEx Consortium (2017). Co-expression networks reveal the tissue-specific regulation of transcription and splicing. Genome Res. *27*, 1843–1858.
83. Li, M.J., Wang, L.Y., Xia, Z., Sham, P.C., and Wang, J. (2013). GWAS3D: Detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. Nucleic Acids Res. *41*, W150-8.
84. Sané, A.T., Seidman, E., Peretti, N., Kleme, M.L., Delvin, E., Deslandres, C., Garofalo, C., Spahis, S., and Levy, E. (2017). Understanding chylomicron retention disease through Sar1b Gtpase gene disruption: insight from cell culture. Arterioscler. Thromb. Vasc. Biol. *37*, 2243–2251.
85. Yuasa, K., Matsuda, T., and Tsuji, A. (2011). Functional regulation of transient receptor potential canonical 7 by cGMP-dependent protein kinase Iα. Cell. Signal. *23*, 1179–1187.
86. Ilegems, E., Iwatsuki, K., Kokrashvili, Z., Benard, O., Ninomiya, Y., and Margolskee, R.F. (2010). REEP2 enhances sweet receptor function by recruitment to lipid rafts. J. Neurosci. *30*, 13774–13783.

87. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.
88. Pattin, K.A., and Moore, J.H. (2009). Role for protein-protein interaction databases in human genetics. Expert Rev. Proteomics *6*, 647–659.
89. Taniguchi, K., Matsumura, K., Kageyama, S., Ii, H., Ashihara, E., Chano, T., Kawauchi, A., Yoshiki, T., and Nakata, S. (2018). Prohibitin-2 is a novel regulator of $p21^{WAF1/CIP1}$ induced by depletion of γ-glutamylcyclotransferase. Biochem. Biophys. Res. Commun. *496*, 218–224.
90. Fimia, G.M., Corazzari, M., Antonioli, M., and Piacentini, M. (2013). Ambra1 at the crossroad between autophagy and cell death. Oncogene *32*, 3311–3318.
91. Fimia, G.M., Stoykova, A., Romagnoli, A., Giunta, L., Di Bartolomeo, S., Nardacci, R., Corazzari, M., Fuoco, C., Ucar, A., Schwartz, P., et al. (2007). Ambra1 regulates autophagy and development of the nervous system. Nature *447*, 1121–1125.
92. Frazier, A.E., Thorburn, D.R., and Compton, A.G. (2017). Mitochondrial energy generation disorders: genes, mechanisms and clues to pathology. J. Biol. Chem., jbc.R117.809194.
93. Malik, A.N., Parsade, C.K., Ajaz, S., Crosby-Nwaobi, R., Gnudi, L., Czajka, A., and Sivaprasad, S. (2015). Altered circulating mitochondrial DNA and increased inflammation in patients with diabetic retinopathy. Diabetes Res. Clin. Pract. *110*, 257–265.
94. Guo, H., Bueler, S.A., and Rubinstein, J.L. (2017). Atomic model for the dimeric F_O region of mitochondrial ATP synthase. Science *358*, 936–940.
95. Nakamoto, R.K., Baylis Scanlon, J.A., and Al-Shawi, M.K. (2008). The rotary mechanism of the ATP synthase. Arch. Biochem. Biophys. *476*, 43–50.
96. Fish, J., Raule, N., and Attardi, G. (2004). Discovery of a major D-loop replication origin reveals two modes of human mtDNA synthesis. Science *306*, 2098–2101.
97. Taanman, J.W. (1999). The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta *1410*, 103–123.
98. Wang, H.N., Chen, H.D., Chen, K.Y., Xiao, J.F., He, K., Xiang, G.A., and Xie, X. (2014). Highly expressed MT-ND3 positively associated with histological severity of hepatic steatosis. APMIS *122*, 443–451.
99. Fetterman, J.L., Liu, C., Mitchell, G.F., Vasan, R.S., Benjamin, E.J., Vita, J.A., Hamburg, N.M., and Levy, D. (2018). Relations of mitochondrial genetic variants to measures of vascular function. Mitochondrion *40*, 51–57.
100. Brown, D.A., Perry, J.B., Allen, M.E., Sabbah, H.N., Stauffer, B.L., Shaikh, S.R., Cleland, J.G., Colucci, W.S., Butler, J., Voors, A.A., et al. (2017). Expert consensus document: Mitochondrial function as a therapeutic target in heart failure. Nat. Rev. Cardiol. *14*, 238–250.
101. McDermott-Roe, C., Leleu, M., Rowe, G.C., Palygin, O., Bukowy, J.D., Kuo, J., Rech, M., Hermans-Beijnsberger, S., Schaefer, S., Adami, E., et al. (2017). Transcriptome-wide co-expression analysis identifies LRRC2 as a novel mediator of mitochondrial and cardiac function. PLoS ONE *12*, e0170458.
102. Lee, H.S., and Park, T. (2017). Pathway-driven approaches of interaction between oxidative balance and genetic polymorphism on metabolic syndrome. Oxid. Med. Cell. Longev. *2017*, 6873197.
103. Go, M.J., Hwang, J.Y., Kim, Y.J., Hee Oh, J., Kim, Y.J., Heon Kwak, S., Soo Park, K., Lee, J., Kim, B.J., Han, B.G., et al. (2013). New susceptibility loci in MYL2, C12orf51 and OAS1 associated with 1-h plasma glucose as predisposing risk factors for type 2 diabetes in the Korean population. J. Hum. Genet. *58*, 362–365.
104. Kristiansson, K., Perola, M., Tikkanen, E., Kettunen, J., Surakka, I., Havulinna, A.S., Stancáková, A., Barnes, C., Widen, E., Kajantie, E., et al. (2012). Genome-wide screen for metabolic syndrome susceptibility Loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. Circ Cardiovasc Genet *5*, 242–249.
105. MacDonald, M.J., and Fahien, L.A. (1990). Insulin release in pancreatic islets by a glycolytic and a Krebs cycle intermediate: contrasting patterns of glyceraldehyde phosphate and succinate. Arch. Biochem. Biophys. *279*, 104–108.
106. MacDonald, M.J., Fahien, L.A., Brown, L.J., Hasan, N.M., Buss, J.D., and Kendrick, M.A. (2005). Perspective: emerging evidence for signaling roles of mitochondrial anaplerotic products in insulin secretion. Am. J. Physiol. Endocrinol. Metab. *288*, E1–E15.
107. Dukes, I.D., McIntyre, M.S., Mertz, R.J., Philipson, L.H., Roe, M.W., Spencer, B., and Worley, J.F., 3rd. (1994). Dependence on NADH produced during glycolysis for beta-cell glucose signaling. J. Biol. Chem. *269*, 10979–10982.
108. El-Hattab, A.W., Emrick, L.T., Hsu, J.W., Chanprasert, S., Jahoor, F., Scaglia, F., and Craigen, W.J. (2014). Glucose metabolism derangements in adults with the MELAS m.3243A>G mutation. Mitochondrion *18*, 63–69.
109. Nóbrega-Pereira, S., Fernandez-Marcos, P.J., Brioche, T., Gomez-Cabrera, M.C., Salvador-Pascual, A., Flores, J.M., Viña, J., and Serrano, M. (2016). G6PD protects from oxidative damage and improves healthspan in mice. Nat. Commun. *7*, 10894.
110. Cunningham, A.D., Colavin, A., Huang, K.C., and Mochly-Rosen, D. (2017). Coupling between protein stability and catalytic activity determines pathogenicity of G6PD variants. Cell Rep. *18*, 2592–2599.
111. Dörner, G., Mohnike, A., and Steindel, E. (1975). On possible genetic and epigenetic modes of diabetes transmission. Endokrinologie *66*, 225–227.
112. Thomas, F., Balkau, B., Vauzelle-Kervroedan, F., Papoz, L.; and CODIAB-INSERM-ZENECA Study Group (1994). Maternal effect and familial aggregation in NIDDM. The CODIAB Study. Diabetes *43*, 63–67.
113. Lin, R.S., Lee, W.C., Lee, Y.T., Chou, P., and Fu, C.C. (1994). Maternal role in type 2 diabetes mellitus: indirect evidence for a mitochondrial inheritance. Int. J. Epidemiol. *23*, 886–890.
114. Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G., et al.; MAGIC investigators; and GIANT Consortium (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat. Genet. *42*, 579–589.

115. Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H., Hirota, Y., Mori, H., Jonsson, A., Sato, Y., et al. (2008). Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat. Genet. *40*, 1092–1097.
116. Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G., Ng, D.P., Holmkvist, J., Borch-Johnsen, K., Jørgensen, T., et al. (2008). SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat. Genet. *40*, 1098–1102.
117. Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C.D., and Spector, D.L. (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell *107*, 727–738.
118. Lai, Y.K., Lai, N.M., and Lee, S.W. (2017). Glucose-6-phosphate dehydrogenase deficiency and risk of diabetes: a systematic review and meta-analysis. Ann. Hematol. *96*, 839–845.
119. Heymann, A.D., Cohen, Y., and Chodick, G. (2012). Glucose-6-phosphate dehydrogenase deficiency and type 2 diabetes. Diabetes Care *35*, e58.
120. van der Wijst, M.G., and Rots, M.G. (2015). Mitochondrial epigenetics: an overlooked layer of regulation? Trends Genet. *31*, 353–356.
121. Schleinitz, D., Klöting, N., Lindgren, C.M., Breitfeld, J., Dietrich, A., Schön, M.R., Lohmann, T., Dreßler, M., Stumvoll, M., McCarthy, M.I., et al. (2014). Fat depot-specific mRNA expression of novel loci associated with waist-hip ratio. Int. J. Obes. *38*, 120–125.
122. Meyre, D., Delplanque, J., Chèvre, J.C., Lecoeur, C., Lobbens, S., Gallina, S., Durand, E., Vatin, V., Degraeve, F., Proença, C., et al. (2009). Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. Nat. Genet. *41*, 157–159.
123. Bambace, C., Dahlman, I., Arner, P., and Kulyté, A. (2013). NPC1 in human white adipose tissue and obesity. BMC Endocr. Disord. *13*, 5.
124. Jelinek, D., Castillo, J.J., and Garver, W.S. (2013). The C57BL/6J Niemann-Pick C1 mouse model with decreased gene dosage has impaired glucose tolerance independent of body weight. Gene *527*, 65–70.
125. Woś, M., Szczepanowska, J., Pikuła, S., Tylki-Szymańska, A., Zabłocki, K., and Bandorowicz-Pikuła, J. (2016). Mitochondrial dysfunction in fibroblasts derived from patients with Niemann-Pick type C disease. Arch. Biochem. Biophys. *593*, 50–59.
126. Bernhard, F., Landgraf, K., Klöting, N., Berthold, A., Büttner, P., Friebe, D., Kiess, W., Kovacs, P., Blüher, M., and Körner, A. (2013). Functional relevance of genes implicated by obesity genome-wide association study signals for human adipocyte biology. Diabetologia *56*, 311–322.
127. Buzaglo-Azriel, L., Kuperman, Y., Tsoory, M., Zaltsman, Y., Shachnai, L., Zaidman, S.L., Bassat, E., Michailovici, I., Sarver, A., Tzahor, E., et al. (2016). Loss of muscle MTCH2 increases whole-body energy utilization and protects from diet-induced obesity. Cell Rep. *14*, 1602–1610.
128. Li, Q., Wang, L., Tan, W., Peng, Z., Luo, Y., Zhang, Y., Zhang, G., Na, D., Jin, P., Shi, T., et al. (2011). Identification of C1qTNF-related protein 4 as a potential cytokine that stimulates the STAT3 and NF-κB pathways and promotes cell survival in human cancer cells. Cancer Lett. *308*, 203–214.
129. Byerly, M.S., Petersen, P.S., Ramamurthy, S., Seldin, M.M., Lei, X., Provost, E., Wei, Z., Ronnett, G.V., and Wong, G.W. (2014). C1q/TNF-related protein 4 (CTRP4) is a unique secreted protein with two tandem C1q domains that functions in the hypothalamus to modulate food intake and body weight. J. Biol. Chem. *289*, 4055–4069.
130. Sepe, M., Festa, L., Tolino, F., Bellucci, L., Sisto, L., Alfano, D., Ragno, P., Calabrò, V., de Franciscis, V., La Mantia, G., and Pollice, A. (2011). A regulatory mechanism involving TBP-1/Tat-Binding Protein 1 and Akt/PKB in the control of cell proliferation. PLoS ONE *6*, e22800.
131. Hinney, A., Albayrak, O., Antel, J., Volckmar, A.L., Sims, R., Chapman, J., Harold, D., Gerrish, A., Heid, I.M., Winkler, T.W., et al.; GERAD Consortium; IGAP Consortium; and GIANT Consortium (2014). Genetic variation at the CELF1 (CUGBP, elav-like family member 1 gene) locus is genome-wide associated with Alzheimer's disease and obesity. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *165B*, 283–293.
132. Ridler, C. (2016). Obesity: Inheritance via mitochondria. Nat. Rev. Endocrinol. *12*, 497.
133. Semenkovich, C.F. (2017). We know more than we can tell about diabetes and vascular disease: The 2016 Edwin Bierman Award Lecture. Diabetes *66*, 1735–1741.
134. Szendroedi, J., Phielix, E., and Roden, M. (2011). The role of mitochondria in insulin resistance and type 2 diabetes mellitus. Nat. Rev. Endocrinol. *8*, 92–103.
135. West, A.P., and Shadel, G.S. (2017). Mitochondrial DNA in innate immune responses and inflammatory pathology. Nat. Rev. Immunol. *17*, 363–375.
136. Francis, J.W., Turn, R.E., Newman, L.E., Schiavon, C., and Kahn, R.A. (2016). Higher order signaling: ARL2 as regulator of both mitochondrial fusion and microtubule dynamics allows integration of 2 essential cell functions. Small GTPases *7*, 188–196.
137. Castellana, S., Fusilli, C., Mazzoccoli, G., Biagini, T., Capocefalo, D., Carella, M., Vescovi, A.L., and Mazza, T. (2017). High-confidence assessment of functional impact of human mitochondrial non-synonymous genome variations by APOGEE. PLoS Comput. Biol. *13*, e1005628.
138. Samuels, D.C., Carothers, A.D., Horton, R., and Chinnery, P.F. (2006). The power to detect disease associations with mitochondrial DNA haplogroups. Am. J. Hum. Genet. *78*, 713–720.
139. Guo, Y., Lanktree, M.B., Taylor, K.C., Hakonarson, H., Lange, L.A., Keating, B.J.; and IBC 50K SNP array BMI Consortium (2013). Gene-centric meta-analyses of 108 912 individuals confirm known body mass index loci and reveal three novel signals. Hum. Mol. Genet. *22*, 184–201.
140. Middelberg, R.P., Ferreira, M.A., Henders, A.K., Heath, A.C., Madden, P.A., Montgomery, G.W., Martin, N.G., and Whitfield, J.B. (2011). Genetic variants in LPL, OASL and TOMM40/APOE-C1-C2-C4 genes are associated with multiple cardiovascular-related traits. BMC Med. Genet. *12*, 123.
141. Kraja, A.T., Chasman, D.I., North, K.E., Reiner, A.P., Yanek, L.R., Kilpeläinen, T.O., Smith, J.A., Dehghan, A., Dupuis, J., Johnson, A.D., et al.; Cross Consortia Pleiotropy Group; Cohorts for Heart and; Aging Research in Genetic Epidemiology; Genetic Investigation of Anthropometric Traits Consortium; Global Lipids Genetics Consortium; Meta-Analyses of Glucose; Insulin-related

traits Consortium; Global BPgen Consortium; ADIPOGen Consortium; Women's Genome Health Study; and Howard University Family Study (2014). Pleiotropic genes for metabolic syndrome and inflammation. Mol. Genet. Metab. *112*, 317–338.

142. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.

The American Journal of Human Genetics, Volume *104*

Supplemental Data

Associations of Mitochondrial and Nuclear Mitochondrial Variants and Genes with Seven Metabolic Traits

Aldi T. Kraja, Chunyu Liu, Jessica L. Fetterman, Mariaelisa Graff, Christian Theil Have, Charles Gu, Lisa R. Yanek, Mary F. Feitosa, Dan E. Arking, Daniel I. Chasman, Kristin Young, Symen Ligthart, W. David Hill, Stefan Weiss, Jian'an Luan, Franco Giulianini, Ruifang Li-Gao, Fernando P. Hartwig, Shiow J. Lin, Lihua Wang, Tom G. Richardson, Jie Yao, Eliana P. Fernandez, Mohsen Ghanbari, Mary K. Wojczynski, Wen-Jane Lee, Maria Argos, Sebastian M. Armasu, Ruteja A. Barve, Kathleen A. Ryan, Ping An, Thomas J. Baranski, Suzette J. Bielinski, Donald W. Bowden, Ulrich Broeckel, Kaare Christensen, Audrey Y. Chu, Janie Corley, Simon R. Cox, Andre G. Uitterlinden, Fernando Rivadeneira, Cheryl D. Cropp, E. Warwick Daw, Diana van Heemst, Lisa de las Fuentes, He Gao, Ioanna Tzoulaki, Tarunveer S. Ahluwalia, Renée de Mutsert, Leslie S. Emery, A. Mesut Erzurumluoglu, James A. Perry, Mao Fu, Nita G. Forouhi, Zhenglong Gu, Yang Hai, Sarah E. Harris, Gibran Hemani, Steven C. Hunt, Marguerite R. Irvin, Anna E. Jonsson, Anne E. Justice, Nicola D. Kerrison, Nicholas B. Larson, Keng-Hung Lin, Latisha D. Love-Gregory, Rasika A. Mathias, Joseph H. Lee, Matthias Nauck, Raymond Noordam, Ken K. Ong, James Pankow, Amit Patki, Alison Pattie, Astrid Petersmann, Qibin Qi, Rasmus Ribel-Madsen, Rebecca Rohde, Kevin Sandow, Theresia M. Schnurr, Tamar Sofer, John M. Starr, Adele M. Taylor, Alexander Teumer, Nicholas J. Timpson, Hugoline G. de Haan, Yujie Wang, Peter E. Weeke, Christine Williams, Hongsheng Wu, Wei Yang, Donglin Zeng, Daniel R. Witte, Bruce S. Weir, Nicholas J. Wareham, Henrik Vestergaard, Stephen T. Turner, Christian Torp-Pedersen, Evie Stergiakouli, Wayne Huey-Herng Sheu, Frits R. Rosendaal, M. Arfan Ikram, Oscar H. Franco, Paul M. Ridker, Thomas T. Perls, Oluf Pedersen, Ellen A. Nohr, Anne B. Newman, Allan Linneberg, Claudia Langenberg, Tuomas O. Kilpeläinen, Sharon L.R. Kardia, Marit E. Jørgensen, Torben Jørgensen, Thorkild I.A. Sørensen, Georg Homuth, Torben Hansen, Mark O. Goodarzi, Ian J. Deary, Cramer Christensen, Yii-Der Ida Chen, Aravinda Chakravarti, Ivan Brandslund, Klaus Bonnelykke, Kent D. Taylor, James G. Wilson, Santiago Rodriguez, Gail Davies, Bernardo L. Horta, Bharat Thyagarajan, D.C. Rao, Niels Grarup, Victor G. Davila-Roman, Gavin Hudson, Xiuqing Guo, Donna K. Arnett, Caroline Hayward, Dhananjay Vaidya, Dennis O. Mook-Kanamori, Hemant K. Tiwari, Daniel Levy, Ruth J.F. Loos, Abbas Dehghan, Paul Elliott, Afshan N. Malik, Robert A. Scott, Diane M. Becker, Mariza de Andrade, Michael A. Province, James B. Meigs, Jerome I. Rotter, and Kari E. North

Supplement:

Figures, Tables, Methods, References, Acknowledgments and Study Descriptions

Content

Table S4. GWAS Publications Summary Results Used for Identifying Significant MT-nDNA Candidate Genes
Table S5. Domain Terms (in Pathways and Metabolism-MeSH) Enriched for MT-nDNA gene Clusters and their Assigned Abstracts in Percent
Table S6. Functional Annotation of 109 MT-nDNA candidate Genes (**Separate Excel Worksheet**)
Table S7. Thirteen and Fifteen Genes out of 109 MT-nDNA Candidates Associated Respectively with T2D and CVD
Table S8. Several Variants of MT-nDNA Candidates Are eSNV, Affecting Expression of Some of Their Corresponding Genes or Other cis-Located Genes
Table S9. Transcription Regulation Evidence on 588 SNVs of 109 MT-nDNA Genes Using RegulomeDB and HaploReg (**Separate Excel Worksheet**)
Table S10. Transcription Regulation of MT-nDNA Candidate Variants Using GTEx Software (**Separate Excel worksheet**)
Table S11. Fifteen Genes out of 109 MT-nDNA Candidates Ranked with Highest PageRank Score for PPI
Table S12. A Comparison of 109 MT-nDNA Candidate Genes Detected with $P<1E-06$, and Additional SNVs for the Same Genes from GWAS-Catalog Conditional They Pass $P\leq 5E-08$ (**Separate Excel Worksheet**)
Table S13. A Comparison of mtDNA Association Results of Three Papers with Our Overlapping Results for Different Traits (**Separate Excel Worksheet**)
Table S14. Summary of Chips Used for Studying mtDNA Variants per Cohort and Summary Statistics for Imputation
Table S15. Illumina Infinium Omni Express Exome v1.1
Table S16. Infinium Core Exome v1.0
Table S17. Infinium Core Exome v1.1
Table S18. Infinium Human Exome v1.2
Table S19. Illumina HumanExome-12v1-1
Table S20. Illumina Human 1M Duo v3
Table S21. Human Omni2.5-8v1
Table S22. Illumina Human660W-Quad_v1
Table S23. Affymetrix 6.0
Table S24. Illumina HumanHap550v3.0
Table S25. Illumina CardioMetaboChip
Table S26. Illumina Human 610 Quad v1
Table S27. Affymetrix Axiom UKBiobank
Table S28. Illumina Human1Mv1C
Table S29. MT-MeSH Terms Extracted from Literature Lab
Table S30. Selected 2,282 Candidate Genes (MT-nDNA) that May Contribute to Mitochondria (**Separate Excel Worksheet**)
Table S31. Final list of genes (HUGO approved names) as mtDNA- and possible MT-genes candidates for contributing to mitochondria and corresponding MIM number and location. (**Separate Excel Worksheet**)
Table S32. Predicted Functional Effects of mtDNA Non-Synonymous Mutations Present in Our Arrays. (**Separate DOCX Table**)

3. **Supplementary Methods**
    1. Phenotype Preparation
    2. Mitochondrial Single Nucleotide Variant Annotation
    3. mtDNA Genotyping and Imputation QC
    4. mtDNA Variants Association Tests (Single variant linear regression and gene based SKAT tests)
    5. mtDNA Meta-Analyses (single-variant and gene-based)
    6. mtDNA Genomic Associations
    7. MT-nDNA Candidate Genes
    8. Enrichment Analysis of MT-nDNA Candidate Genes
    9. Biological Functional Annotations

4. **References**

5. **CHARGEmtDNA+ Study Acknowledgments**

6. **CHARGEmtDNA+ Working Group Study Descriptions**

1. **Supplementary Figures**

Figure S1. BMI mtDNA Association Tests and QQ-plot Distribution

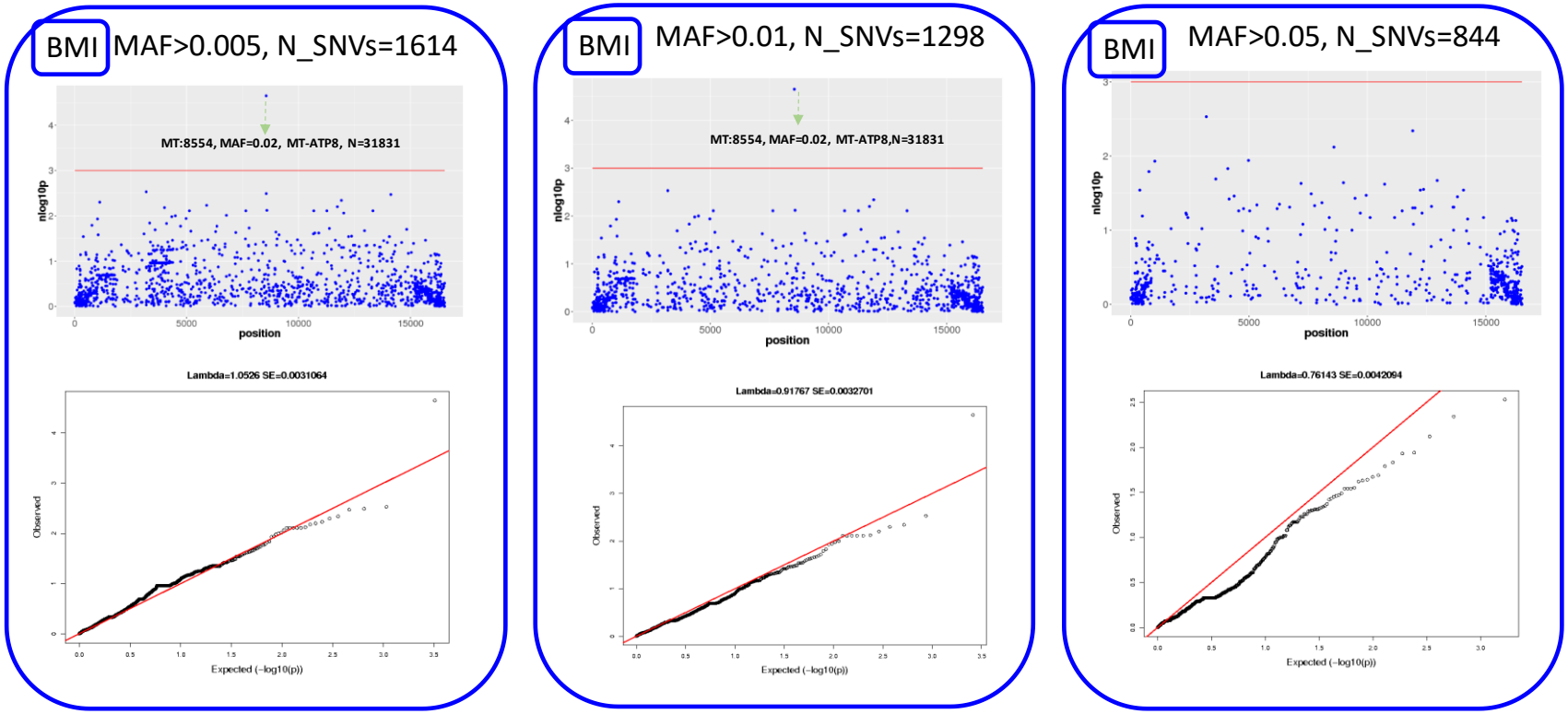


Figure S2. WHR mtDNA Association Tests and QQ-plot Distribution

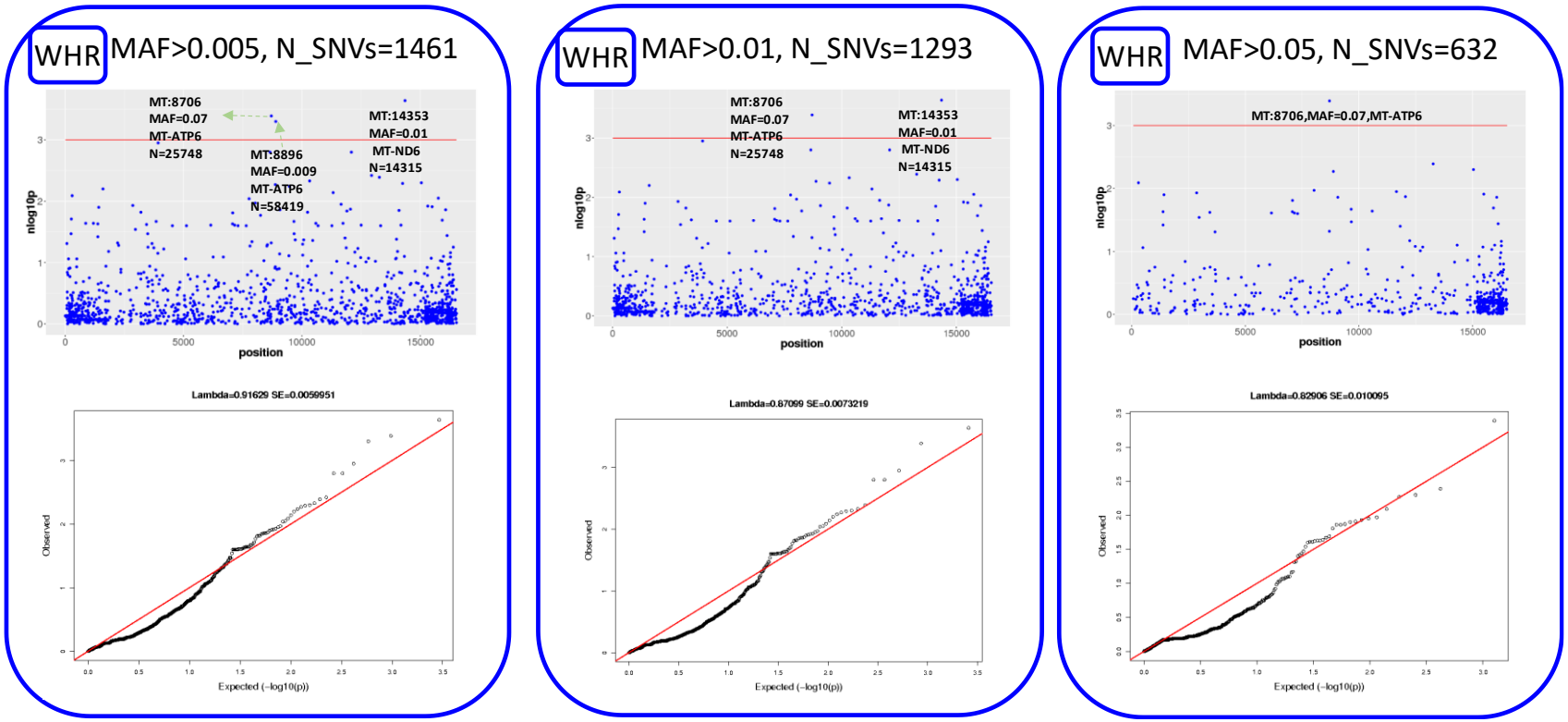


Figure S3. Glucose mtDNA Association Tests and QQ-plot Distribution

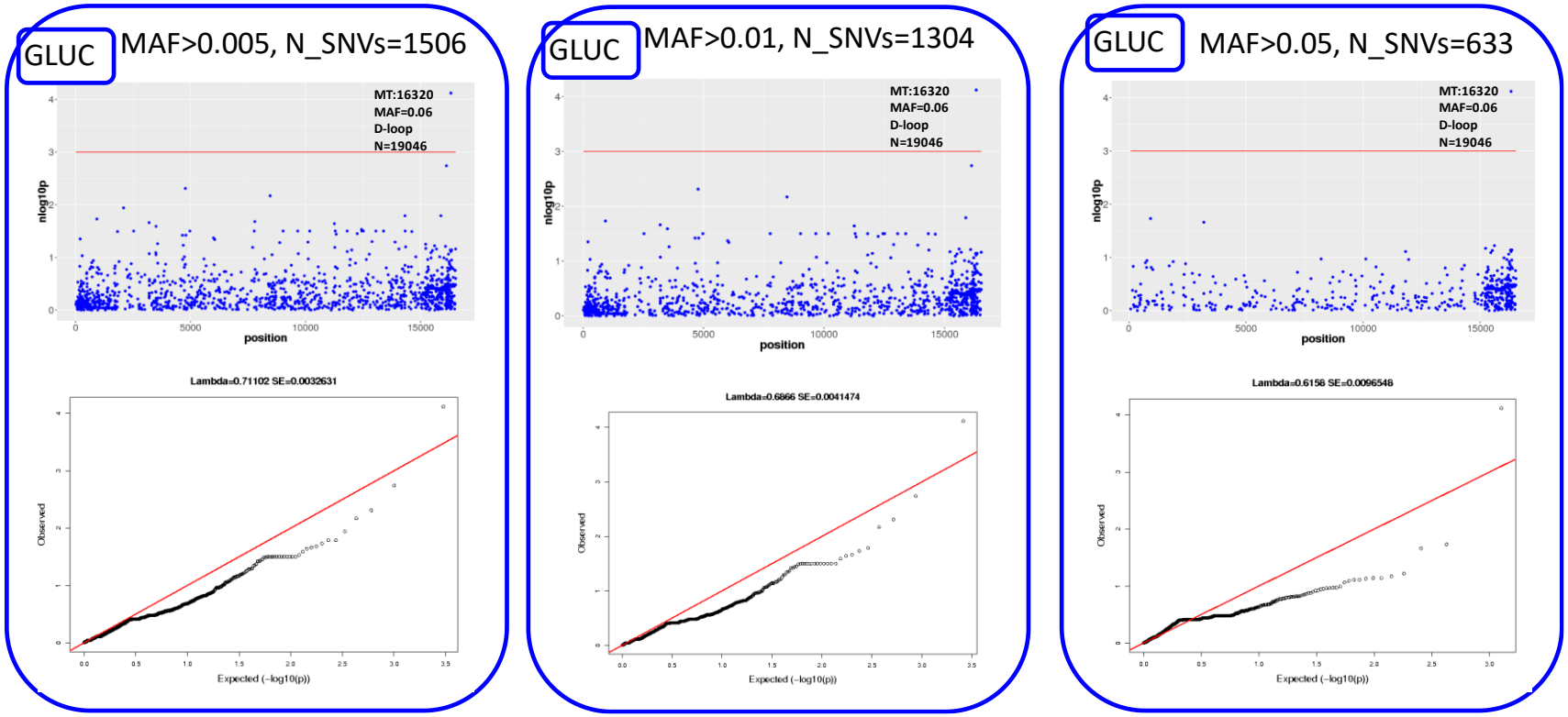


Figure S4. Insulin mtDNA Association Tests and QQ-plot Distribution

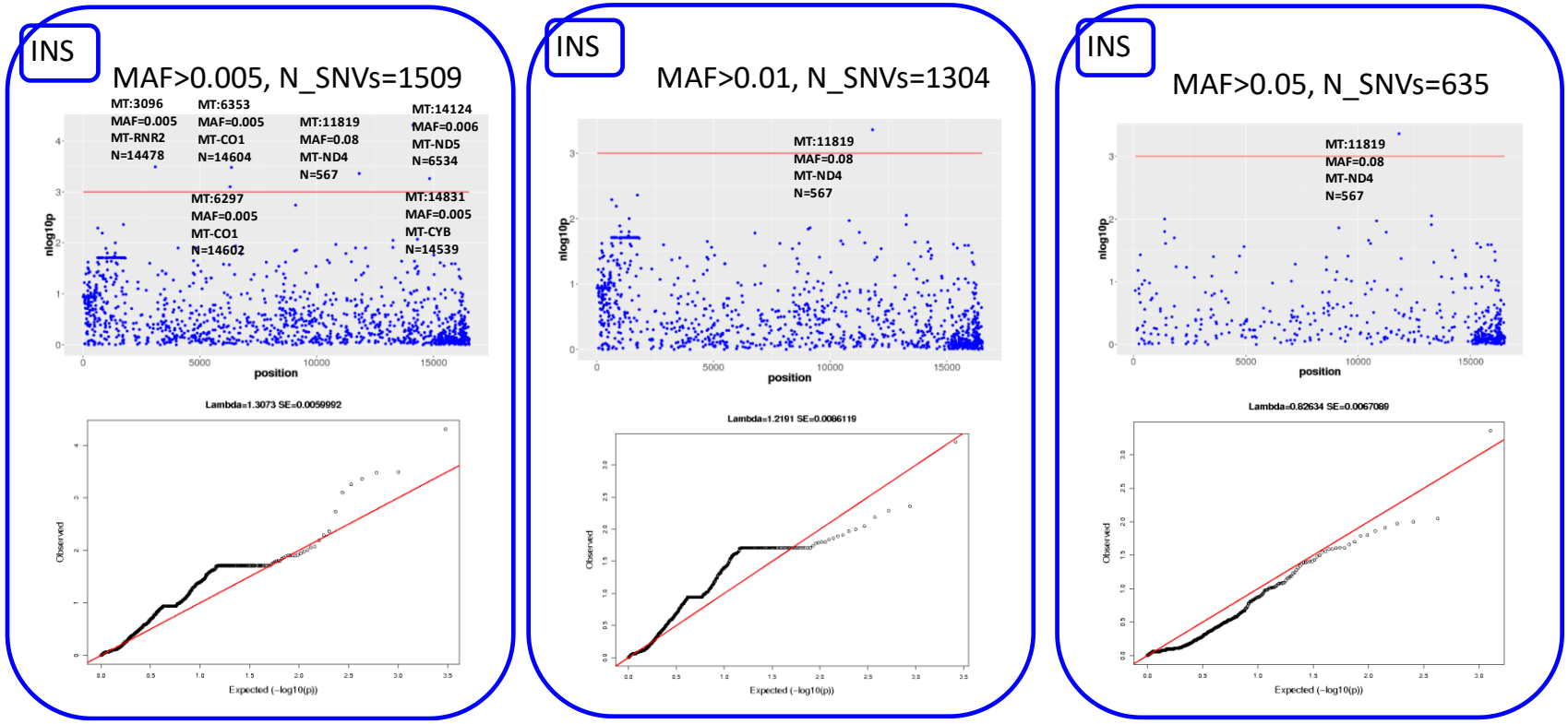


Figure S5. HOMA-B mtDNA Association Tests and QQ-plot Distribution

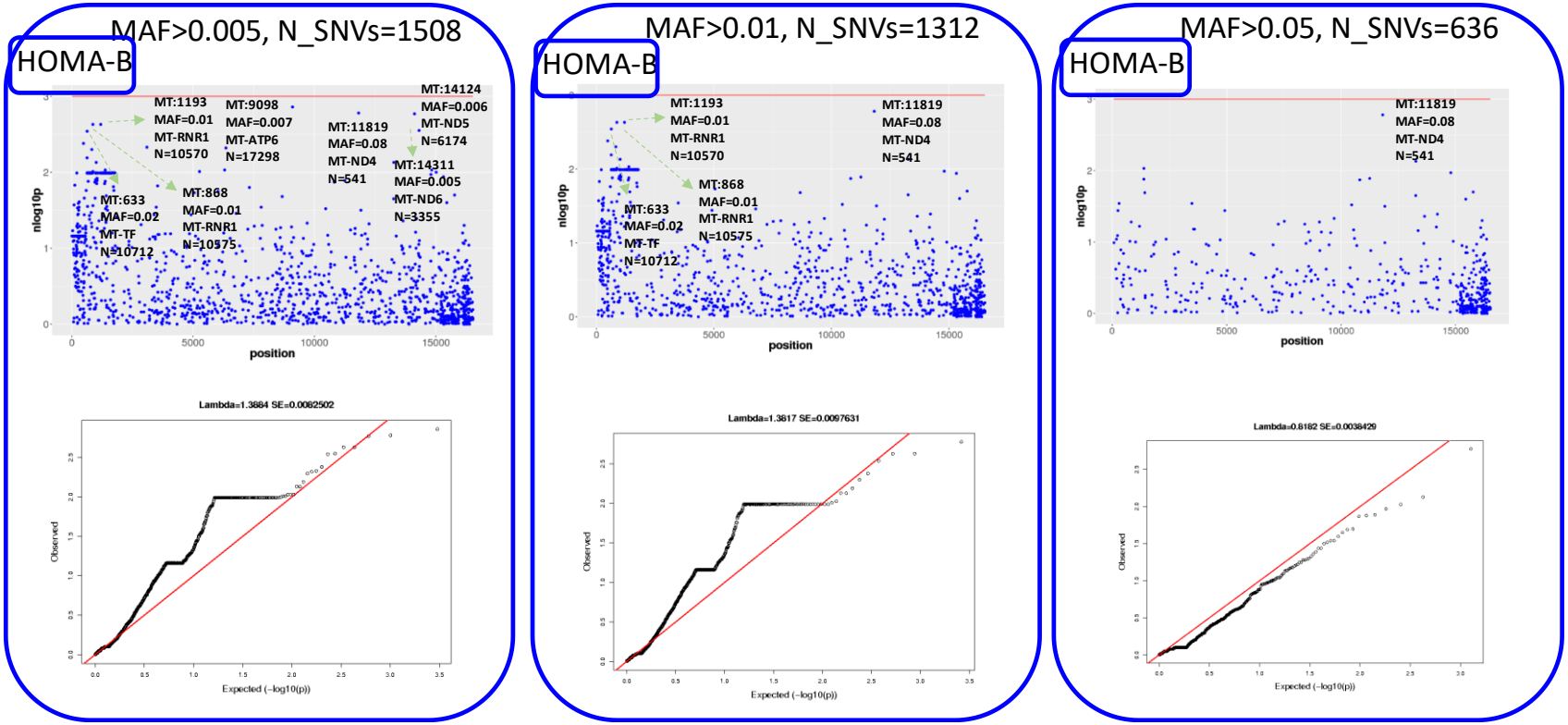


Figure S6. HOMA-IR mtDNA Association Tests and QQ-plot Distribution

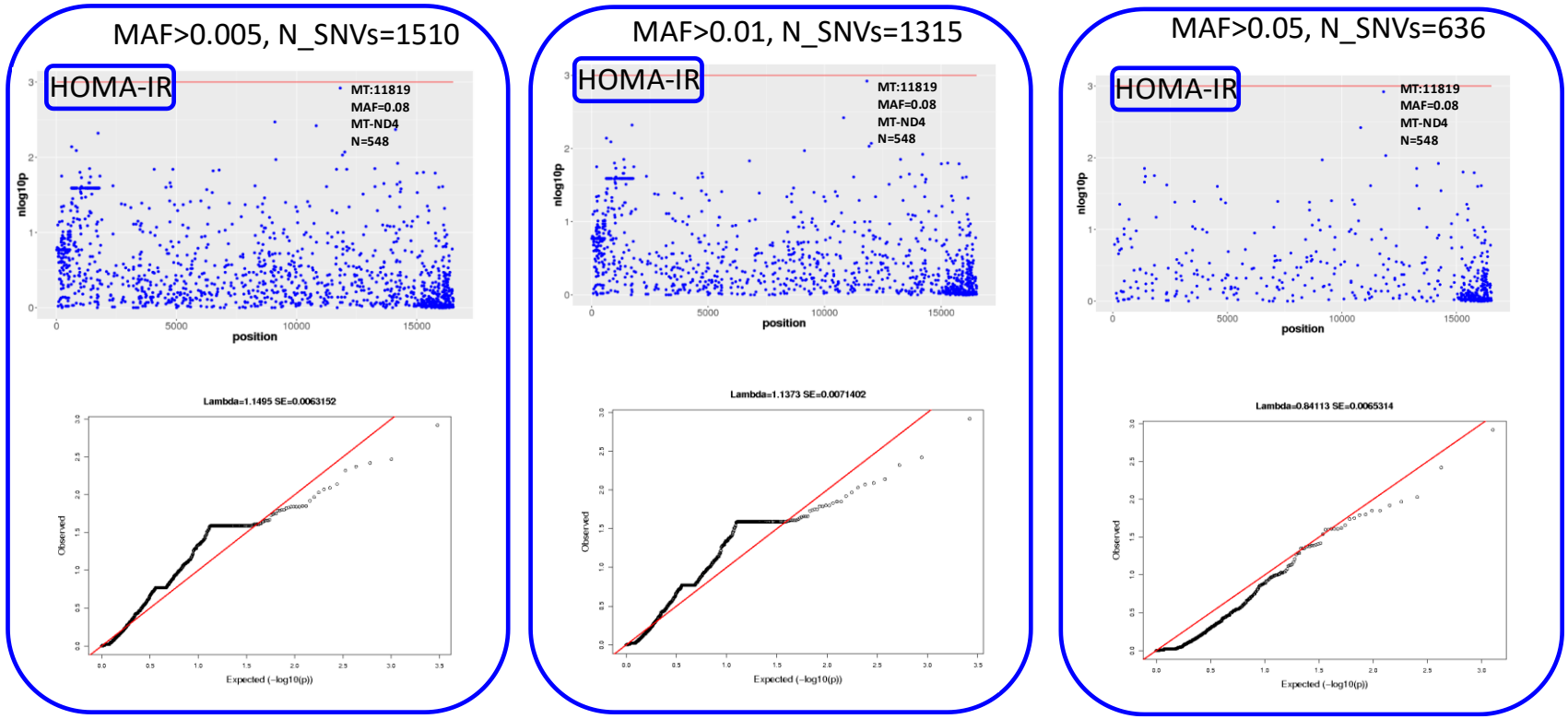


Figure S7. HbA1c mtDNA Association Tests and QQ-plot Distribution

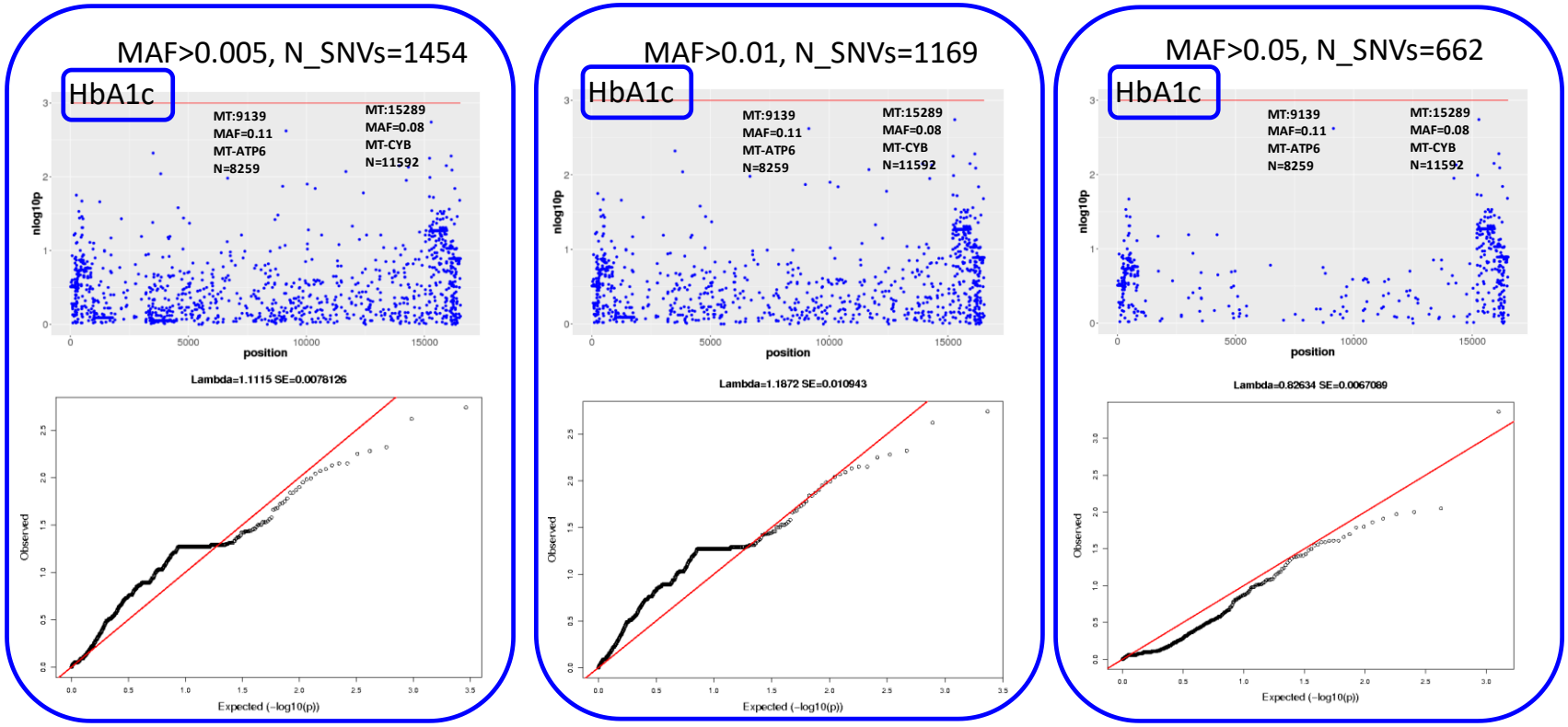


Figure S8. HOMA-B and HOMA-IR Association Results for MT-nDNA Candidate Genes

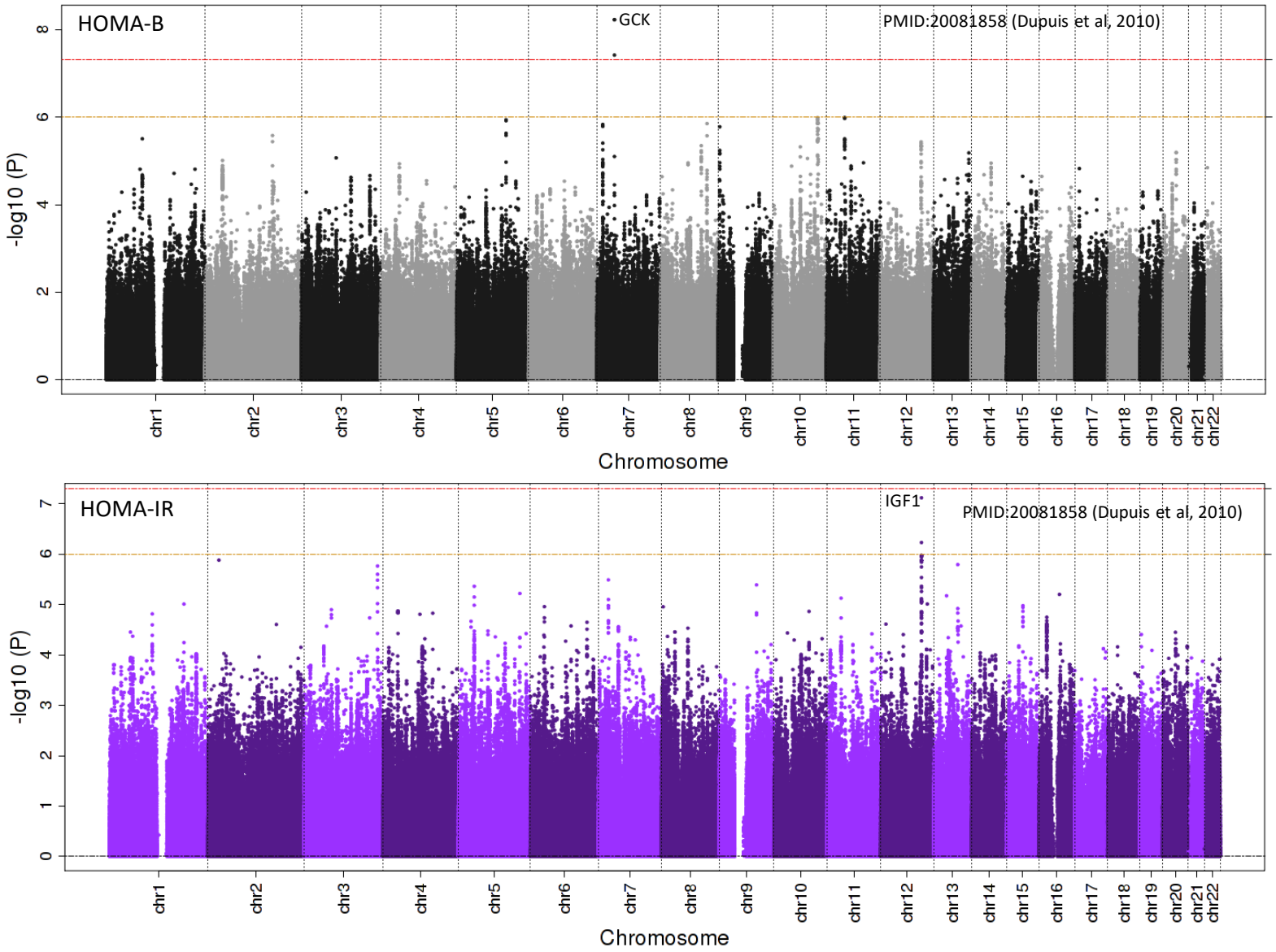


Figure 9. HbA1c Association Results with MT-nDNA Candidate Genes

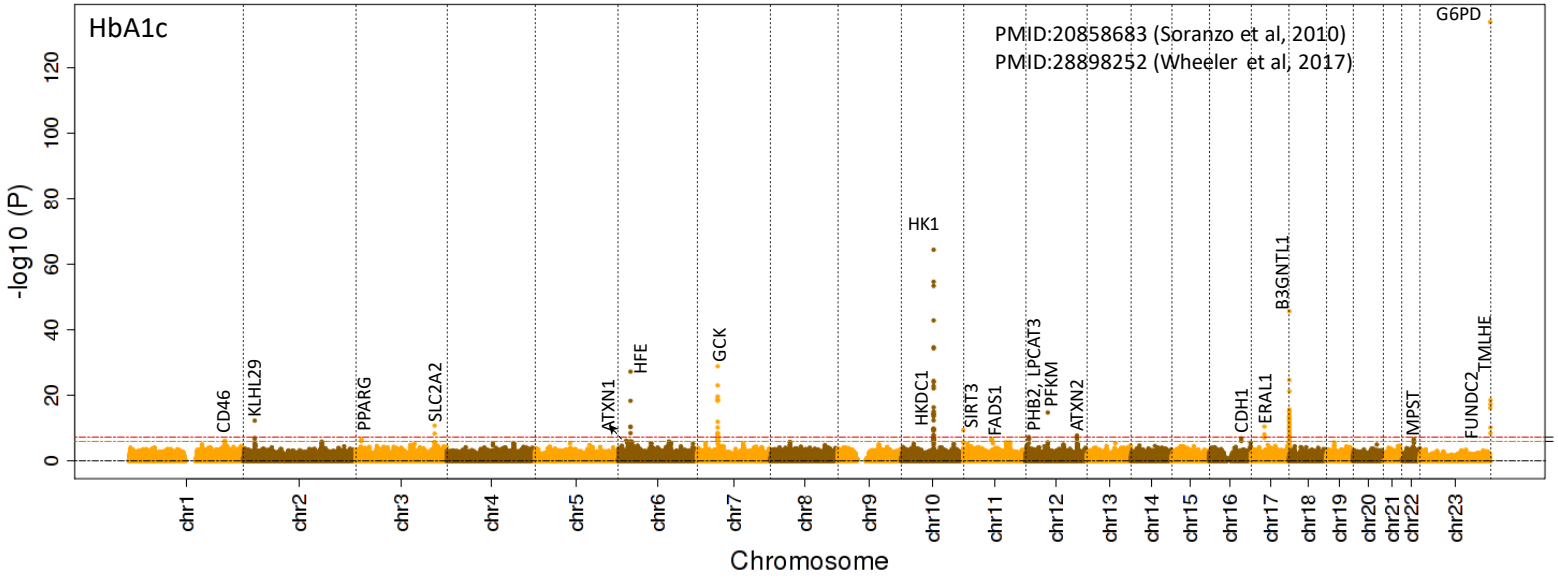


Figure S10. Six Clusters of Associations Among 109 MT-nDNA candidate Genes and Biological Terms Using Literature Lab.

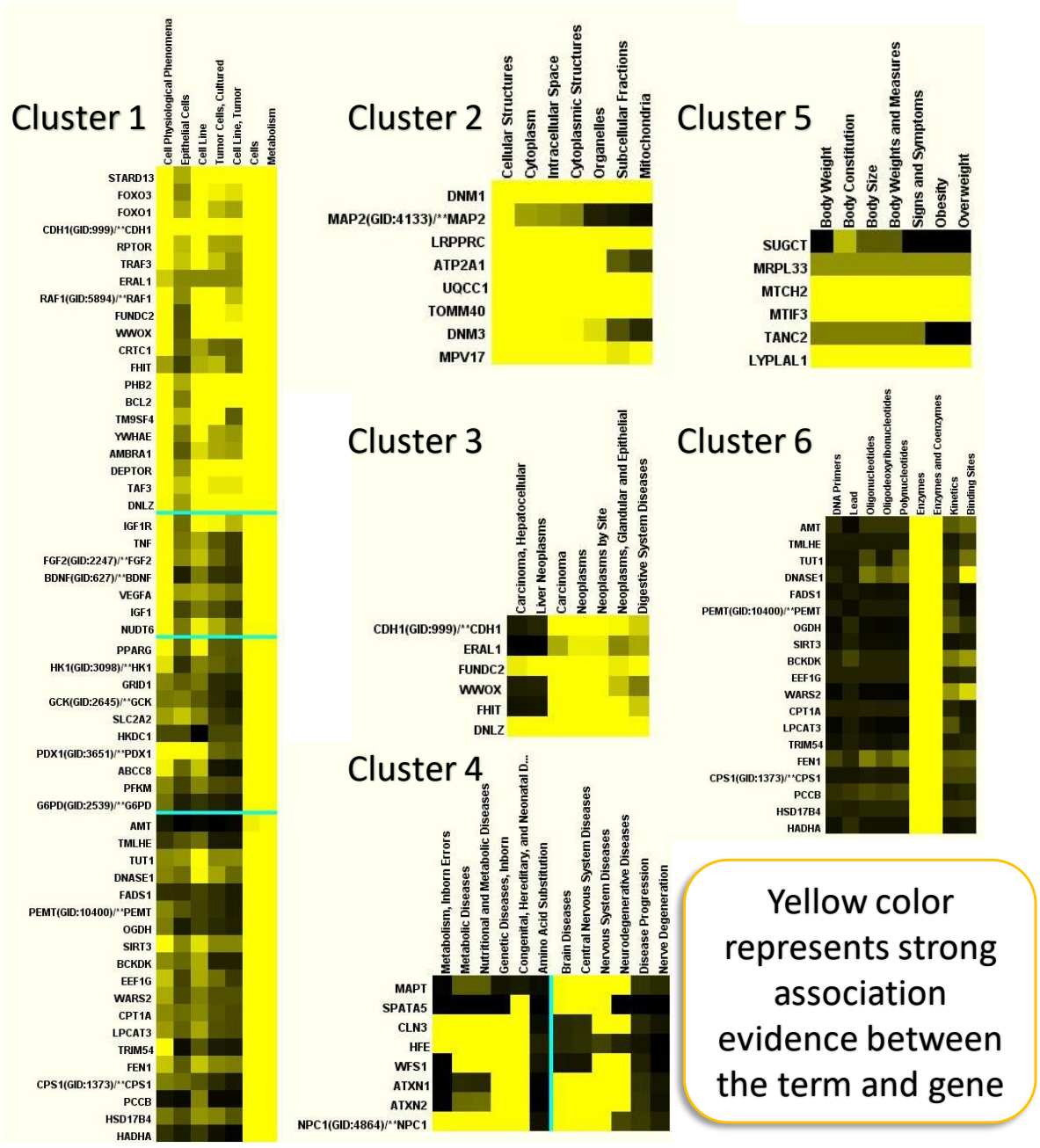


Figure S11. A Diagram on MT-nDNA Candidate Genes Commonalities and Functions Summary. Details are presented in Table 4 and Table S6 and Discussion.

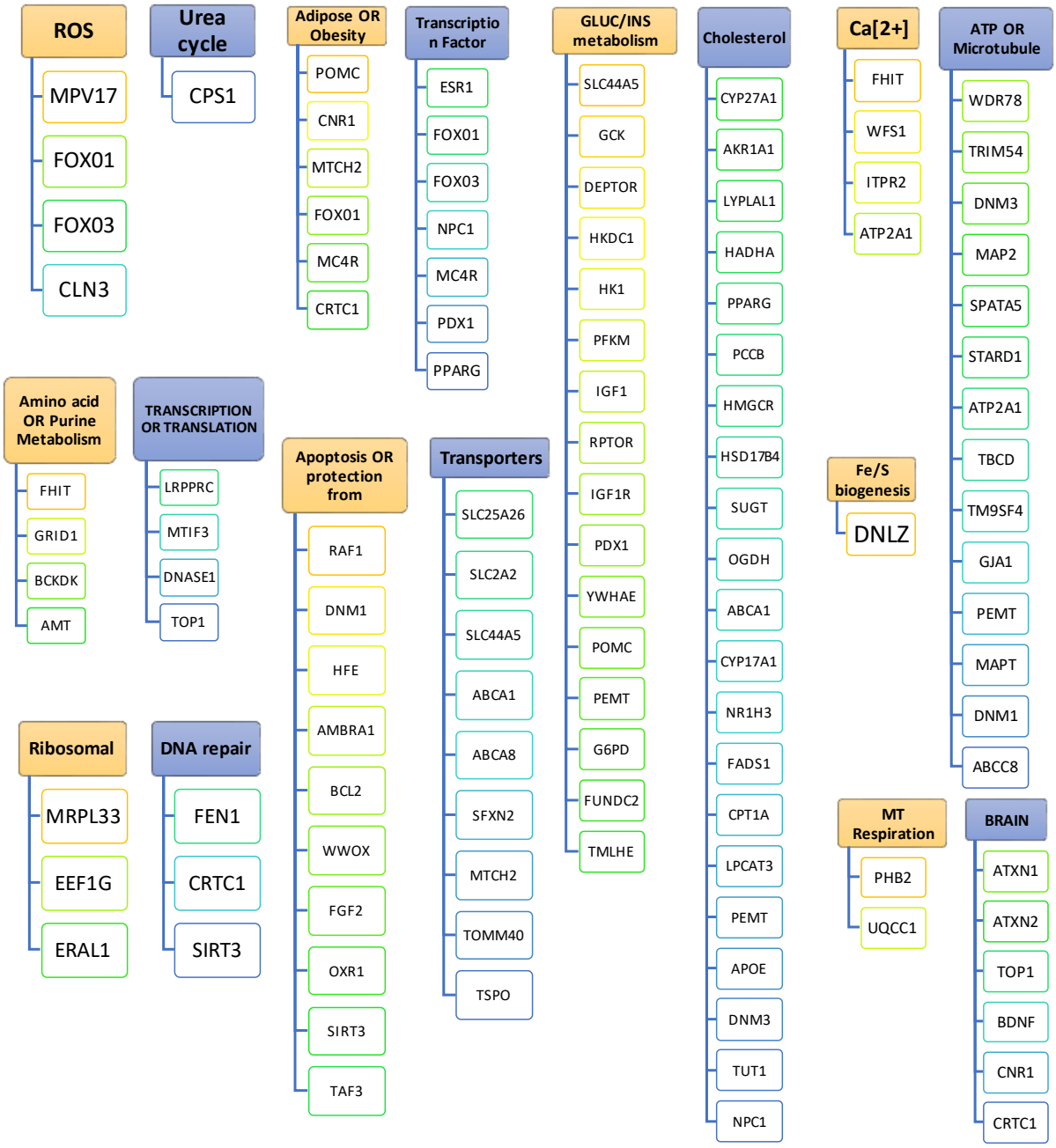
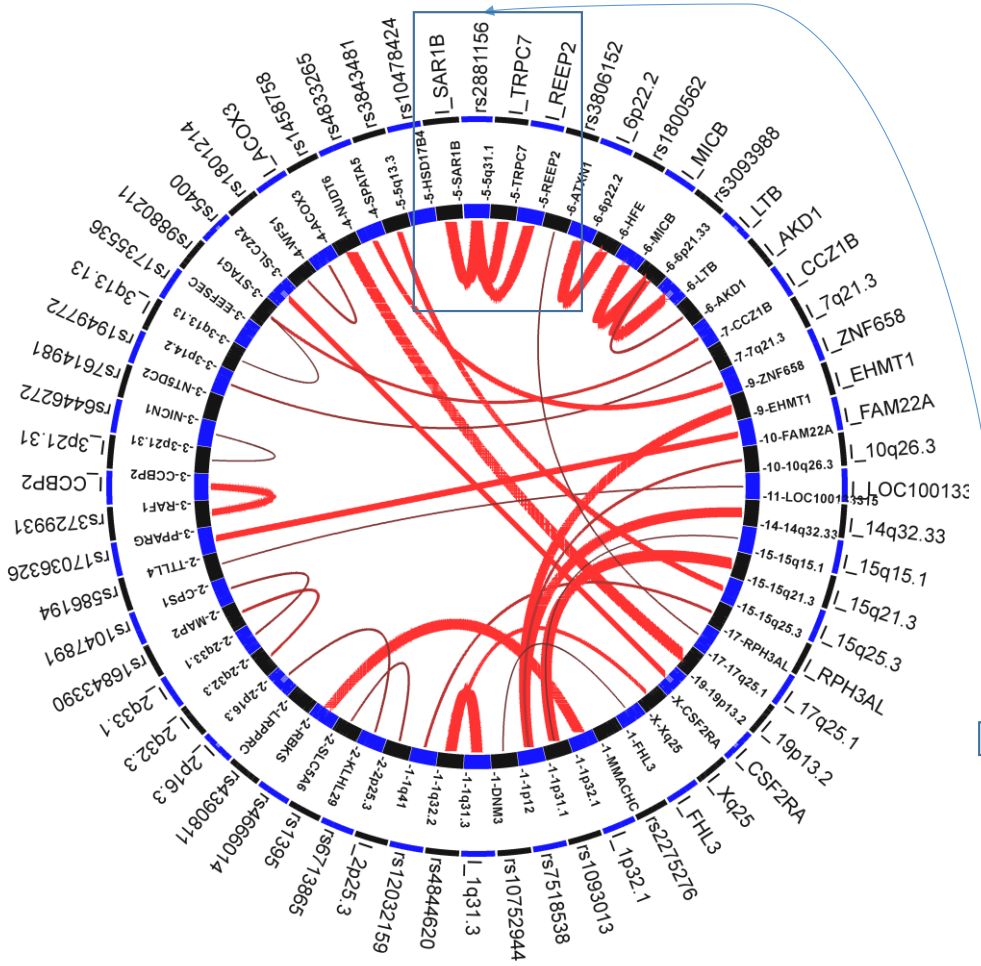


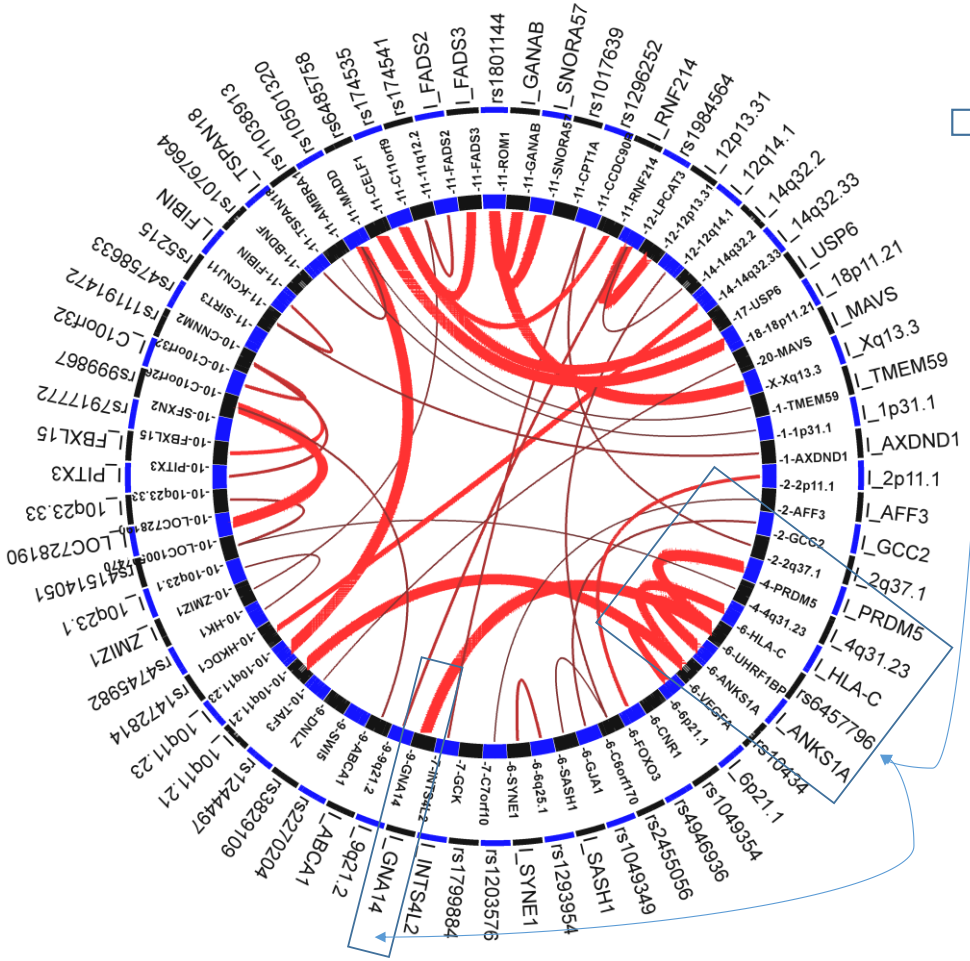
Figure S12.1. Detecting trans-Regulatory Variants by Combining GWAS *P*-value and Probable Functional Variants that Affect Transcriptional Regulation



No	CHRPOS	SNPID	GENOTYPE	LOCUS	FINALP	LeadSNP	LEADSNP_P	RSQUARE
1	1:119750411	rs7518538	A|T	1p12	7.5E-20	rs6428792	8.0E-18	0.857
2	1:172363524	rs10752944	C|T	DNM3	2.4E-21	rs2301453	4.4E-17	0.923
3	1:207980901	rs4844620	G|A	1q32.2	1.1E-09	rs4844390	6.9E-07	0.874
4	1:219586415	rs12032159	G|A	1q41	2.5E-11	rs11118296	3.0E-07	1
5	1:45973928	rs2275276	G|A	MMACHC	1.0E-08	rs622798	8.6E-07	1
6	1:76100343	rs1093013	G|T	1p31.1	4.3E-18	rs1093013	4.9E-17	1
7	2:210519833	rs16843390	A|G	MAP2	8.3E-08	rs16843390	3.3E-08	1
8	2:211540507	rs1047891	C|A	CPS1	2.6E-08	rs715	5.8E-07	0.864
9	2:219610433	rs586194	A|G	TTLL4	1.6E-10	rs933994	8.7E-07	0.92
10	2:23899807	rs6713865	A|G	KLHL29	5.2E-15	rs6713865	4.4E-13	1
11	2:25389224	rs934778	A|G	POMC	3.9E-06	rs934778	7.2E-07	1
12	2:26458244	rs1107238	T|C	HADHA	4.1E-07	rs1107238	8.9E-07	1
13	2:27424636	rs1395	G|A	SLC5A6	1.2E-12	rs4665965	1.8E-09	0.908
14	2:28019175	rs4666014	G|A	RBKS	3.8E-15	rs3736594	3.0E-13	0.804
15	2:44176225	rs4390811	G|C	LRPPRC	5.4E-09	rs4346434	8.5E-08	0.917
16	3:12389313	rs17036326	A|G	PPARG	1.4E-15	rs17036328	3.6E-12	1
17	3:12626516	rs3729931	G|A	RAF1	5.8E-14	rs3729931	3.6E-10	1
18	3:128096613	rs1735536	G|C	EEFSEC	6.5E-10	rs1735536	3.0E-10	1
19	3:136107549	rs9880211	G|A	STAG1	9.0E-10	rs9844666	7.2E-07	1
20	3:170732300	rs5400	G|A	SLC2A2	1.0E-19	rs11924648	1.0E-17	0.926
21	3:49463287	rs6446272	G|A	NICN1	4.8E-11	rs11715915	4.9E-08	0.944
22	3:52566914	rs7614981	C|A	NT5DC2	7.9E-12	rs12489828	2.6E-10	1
23	3:61250437	rs1949772	G|A	3p14.2	8.0E-19	rs2365389	3.8E-15	0.958
24	3:66373455	rs6782968	A|G	SLC25A26	1.8E-07	rs332375	7.3E-07	1
25	4:123762897	rs10518406	A|G	FGF2	7.9E-07	rs10518406	6.5E-07	1
26	4:123835879	rs1458758	T|G	NUDT6	4.9E-10	rs1458758	6.2E-08	1
27	4:124054807	rs4833265	C|T	SPATA5	2.6E-09	rs303084	3.4E-07	1
28	4:6303022	rs1801214	C|A,G,T	WFS1	7.5E-12	rs10012946	4.2E-07	1
29	5:118789020	rs10478424	A|T	HSD17B4	1.6E-08	rs10478424	1.4E-07	1
30	5:135148662	rs2881156	A|G	5q31.1	3.2E-09	rs2881156	8.1E-07	1
31	5:74628375	rs3843481	A|T	5q13.3	3.0E-11	rs12654264	1.8E-08	1
32	6:16746624	rs3806152	G|A	ATXN1	2.7E-09	rs3828870	5.5E-07	1
33	6:26093141	rs1800562	G|A	HFE	1.3E-28	rs1800562	4.7E-28	1
34	6:31492453	rs3093988	G|A	6p21.33	1.3E-08	rs1800629	7.3E-07	0.894

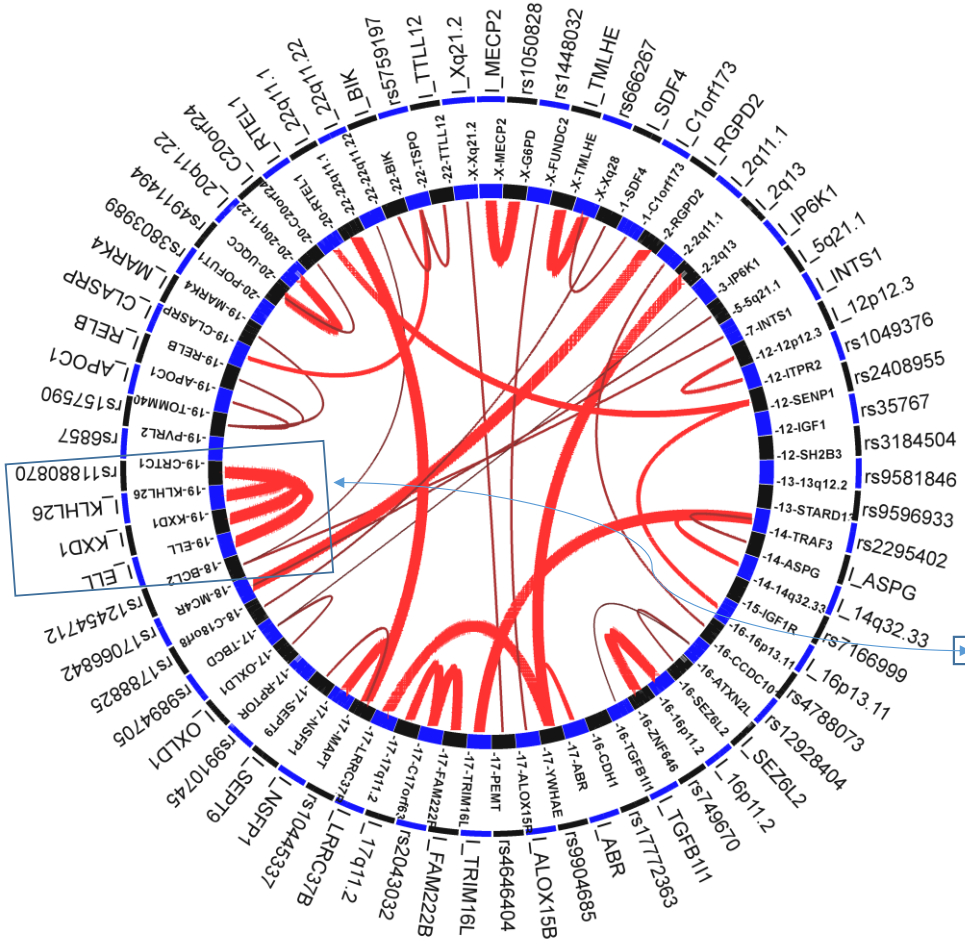
Footnotes: The red line thickness depicts the significance (FINALP in the corresponding tables embedded into Figures S12.1-S12.3). An eSNV of our study (or an eSNV proxy in very high LD with ours) in the outer circle connects with a gene(s)/ cytogenetic band(s), where the "I" letter in front of it means trans-regulatory Interaction (see Supplement, section 8).

Figure S12.2. Detecting trans-Regulatory Variants by Combining GWAS *P*-value and Probable Functional Variants that Affect Transcriptional Regulation



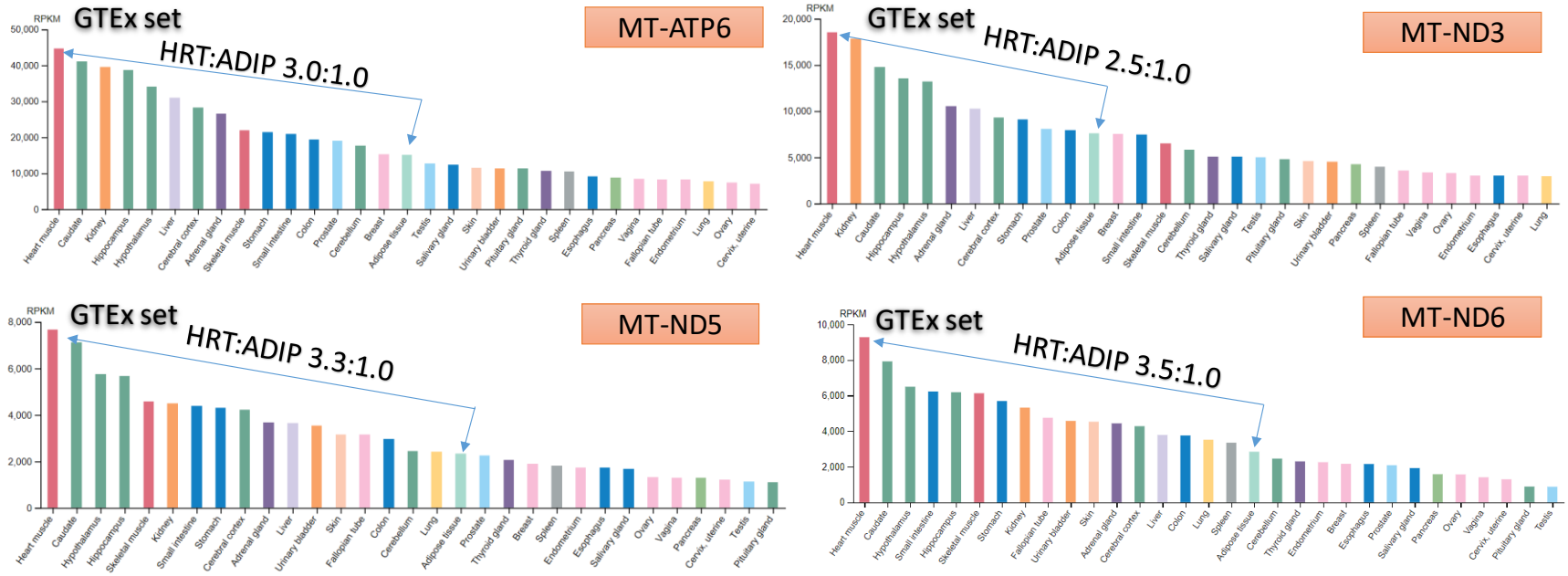
No	CHRPOS	SNPID	GENOTYPE	LOCUS	FINALP	LeadSNP	LEADSNP_P	RSQUARE
1	6:34828553	rs6457796	T|C	UHRF1BP1	1.0E-11	rs6457796	1.2E-09	1
2	6:43753212	rs10434	A|G	VEGFA	1.4E-08	rs10434	8.8E-07	1
3	6:88853451	rs1049354	A|G	CNR1	3.2E-08	rs1049354	9.6E-07	1
4	6:109003321	rs4946936	T|C	FOXO3	1.8E-10	rs9400239	1.6E-08	0.908
5	6:121482620	rs2455056	T|A	C6orf170	1.5E-14	rs1273733	4.0E-12	1
6	6:121770642	rs1049349	A|T	GJA1	2.9E-17	rs1049349	4.2E-15	1
7	6:151990961	rs1293954	A|G	6q25.1	4.5E-09	rs1293954	4.4E-09	1
8	7:40847832	rs1203576	A|T	C7orf10	8.6E-11	rs1203576	1.5E-10	1
9	7:44229068	rs1799884	C|T	GCK	1.2E-90	rs2908289	3.3E-88	1
10	7:44744813	rs1088867	A|G	OGDH	2.1E-07	rs1088867	7.9E-08	1
11	8:107511933	rs16892421	T|G	OXR1	1.5E-06	rs16892421	6.9E-07	1
12	8:121042452	rs7835803	A|G	DEPTOR	8.0E-07	rs7835803	5.0E-07	1
13	9:107672365	rs2777795	G|A	ABCA1	9.3E-08	rs2777795	3.1E-08	1
14	9:131042734	rs2270204	T|G	SWI5	3.8E-08	rs7023913	7.3E-07	1
15	9:139256766	rs3829109	G|A	DNLZ	4.6E-10	rs3829109	1.1E-10	1
16	10:7879982	rs1244497	A|G	TAF3	2.9E-12	rs1244497	1.8E-11	1
17	10:71003612	rs1472814	A|G	HKDC1	4.3E-15	rs5030913	3.6E-13	0.955
18	10:71089843	rs4745982	T|G	HK1	4.5E-63	rs4745982	2.9E-65	1
19	10:87359055	rs41514051	A|G	LOC100507470	2.0E-10	rs7899106	3.0E-08	0.824
20	10:104487443	rs7917772	G|A	SFXN2	3.8E-09	rs7917772	1.5E-09	1
21	10:104504564	rs999867	C|T	C10orf26	1.7E-08	rs1004467	1.2E-07	0.866
22	10:104707016	rs11191472	A|T	CNNM2	3.7E-11	rs3740390	4.8E-08	1
23	11:219538	rs4758633	A|G	SIRT3	2.4E-10	rs4758633	3.4E-10	1
24	11:17408630	rs5215	C|T	KCNJ11	5.9E-09	rs757110	4.2E-07	0.921
25	11:27725986	rs10767664	T|A	BDNF	1.1E-16	rs10767664	5.5E-13	1
26	11:46559730	rs11038913	T|C	AMBRA1	9.8E-20	rs11038913	4.9E-18	1
27	11:47293799	rs10501320	G|C	MADD	5.5E-47	rs11039149	4.1E-45	0.908
28	11:47530024	rs6485758	G|A	CELF1	8.2E-31	rs7118178	2.2E-29	0.865
29	11:61551356	rs174535	T|C	C11orf9	1.1E-20	rs174556	7.8E-18	0.808
30	11:61565908	rs174541	T|C	11q12.2	1.1E-10	rs4246215	4.5E-11	1
31	11:62381808	rs1801144	G|C	ROM1	1.3E-12	rs7943191	4.4E-08	0.916
32	11:68598534	rs1017639	A|C	CPT1A	1.9E-10	rs1017639	5.0E-10	1
33	11:82984311	rs1296252	G|T	CCDC90B	6.1E-08	rs1296252	7.6E-08	1
34	12:7090193	rs1984564	A|G	LPCAT3	3.0E-11	rs2110073	4.4E-08	1

Figure S12.3. Detecting trans-Regulatory Variants by Combining GWAS *P*-value and Probable Functional Variants that Affect Transcriptional Regulation



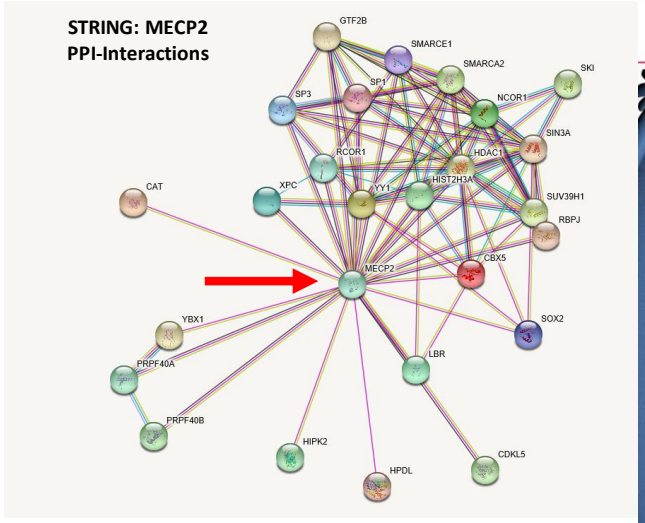
No	CHRPOS	SNPID	GENOTYPE	LOCUS	FINALP	LeadSNP	LEADSNP_P	RSQUARE
1	12:102875569	rs35767	A|G	IGF1	5.9E-09	rs35767	7.3E-08	1
2	12:111884608	rs3184504	T|C	SH2B3	2.2E-08	rs4766578	2.9E-07	0.96
3	12:26491475	rs1049376	T|C	ITPR2	6.1E-17	rs1049380	1.4E-13	1
4	12:48499131	rs2408955	G|T	SENP1	5.5E-16	rs2408955	1.4E-15	1
5	13:27987953	rs9581846	A|G	13q12.2	6.3E-11	rs9581856	1.0E-08	1
6	13:28479963	rs17826899	A|G	13q12.2	2.6E-09	rs1124607	4.5E-07	1
7	13:33767700	rs9596933	G|A	STARD13	6.2E-10	rs1325363	7.3E-08	1
8	13:41138020	rs1078892	T|C	FOXO1	4.6E-07	rs1078892	5.1E-08	1
9	14:103338324	rs2295402	T|A	TRAF3	1.4E-09	rs7143963	2.8E-07	0.862
10	15:99292805	rs7166999	A|C	IGF1R	1.7E-09	rs12908437	6.3E-07	0.953
11	16:28594549	rs4788073	A|G	CCDC101	2.9E-10	rs151181	2.1E-07	0.873
12	16:28847246	rs12928404	T|C	ATXN2L	2.0E-20	rs8055138	8.2E-17	0.869
13	16:31088625	rs749670	A|G	ZNF646	8.8E-13	rs749767	1.2E-09	1
14	16:3711656	rs7200737	G|A	TRAP1	2.0E-07	rs740862	3.2E-07	0.941
15	16:68788909	rs17772363	G|T	CDH1	1.6E-10	rs7186084	1.1E-07	0.83
16	16:78942810	rs1847591	A|G	WWOX	3.0E-06	rs1847591	1.0E-06	1
17	17:1255395	rs9904685	T|C	YWHAE	8.5E-09	rs9904685	6.3E-07	1
18	17:17420199	rs4646404	G|A	PEMT	2.8E-10	rs4646404	5.3E-11	1
19	17:27136692	rs2043032	A|G	C17orf63	6.4E-15	rs9914988	2.8E-11	0.941
20	17:44067400	rs10445337	T|C	MAPT	6.4E-11	rs242559	8.3E-07	0.943
21	17:61183640	rs1319247	G|A	TANC2	1.2E-07	rs1319247	6.0E-08	1
22	17:78610543	rs9910745	C|T	RPTOR	3.1E-11	rs12940622	2.5E-09	0.91
23	17:80823758	rs9894705	G|T	TBCD	3.9E-47	rs1044661	1.7E-46	0.89
24	18:21109250	rs1788825	T|C,G	C18orf8	1.7E-10	rs1788785	2.0E-08	1
25	18:58040624	rs17066842	G|A	MC4R	2.7E-13	rs17066842	6.4E-14	1
26	18:60845884	rs12454712	T|C	BCL2	2.0E-11	rs12454712	1.1E-09	1
27	19:18830704	rs11880870	A|G	CRTC1	1.4E-10	rs757318	8.8E-09	1
28	19:45392254	rs6857	C|T	PVRL2	4.0E-09	rs2075650	1.3E-08	0.862
29	19:45398716	rs157590	A|C	TOMM40	1.0E-09	rs405509	2.7E-07	0.846
30	20:30809216	rs3803989	G|A	POFUT1	4.5E-09	rs2281361	9.8E-07	0.849
31	20:33971914	rs4911494	C|T	UQCC	1.2E-14	rs878639	1.5E-11	0.959
32	20:39721373	rs2076574	A|G	TOP1	3.8E-07	rs2076574	4.1E-07	1
33	22:37425031	rs5750373	G|A	MPST	2.8E-07	rs5750373	2.2E-07	1
34	22:43550149	rs5759197	T|C	TSPO	8.8E-08	rs2284099	6.7E-07	1
35	X:153764217	rs1050828	C|T	G6PD	1.5E-133	rs1050828	8.2E-135	1
36	X:154280805	rs1448032	G|A	FUNDC2	1.2E-18	rs1448032	5.9E-17	1
37	X:154856709	rs666267	C|G	Xq28	1.8E-21	rs5940514	2.2E-19	0.918

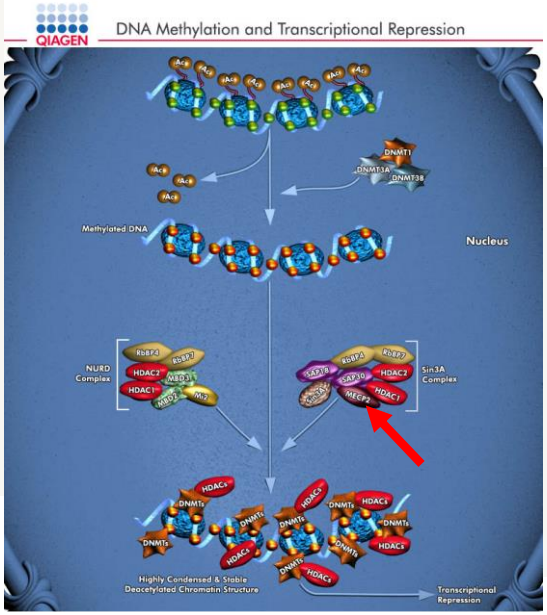
Figure S13. RNA Expression Patterns for Selected mtDNA Genes in Different Human Tissues.



Footnotes: The RNAseq of GTEx is reported as median RPKM (reads per kilobase per million mapped reads) with varying N: 412 for heart muscle to 577 individuals for adipose tissue. The ratio of RNA expression between heart (HRT) and adipose (ADIP) for each gene reflects the trend of more- versus less energy demanding tissues (using tools of Human Protein Atlas (www.proteinatlas.org).

Figure S14. *G6PD* interactions. *G6PD* (chrX:154,531,390-154,547,586, right-picture) trans-regulates *MECP2* (chrX:154,021,573-154,137,103, left picture) and *MECP2* interacts with *SIN3A* (protein-protein interaction evidence), which together (SIN3A and MECP2) contribute to transcription repression of other genes (middle picture)





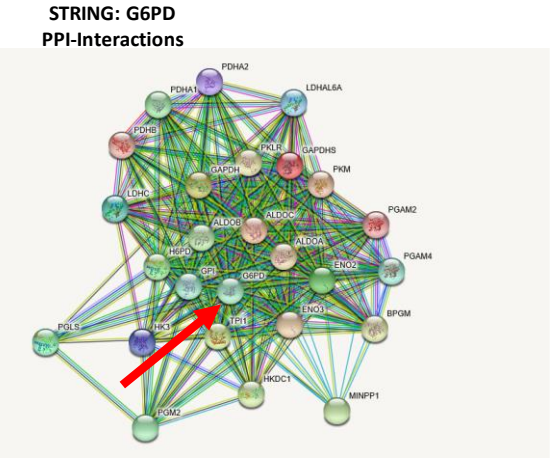


Figure S15. Selection of 25th Upper Quartile LPF of Genes Associated with the Terms "Mitochondrial diseases" and "Pathway analysis".

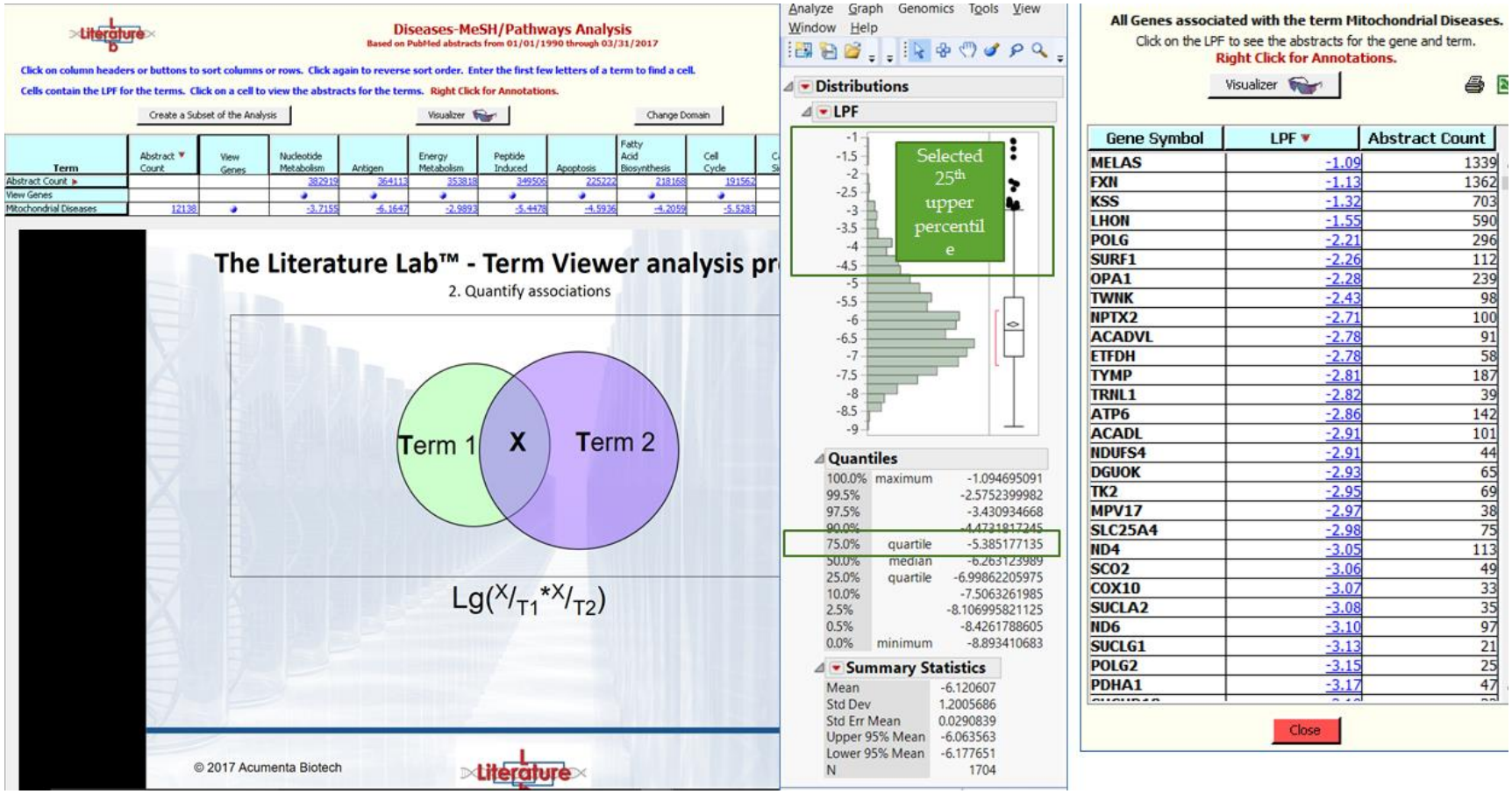
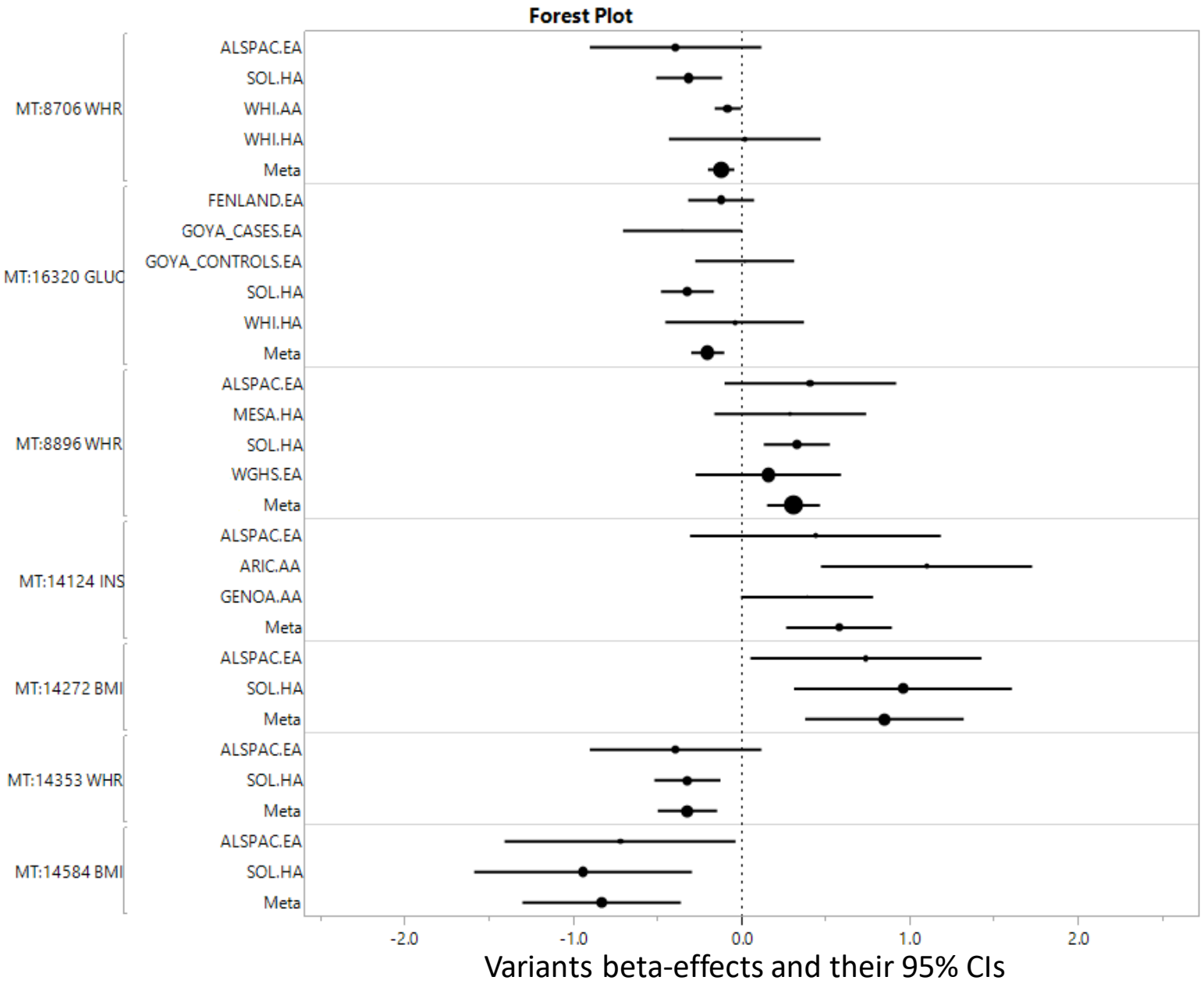
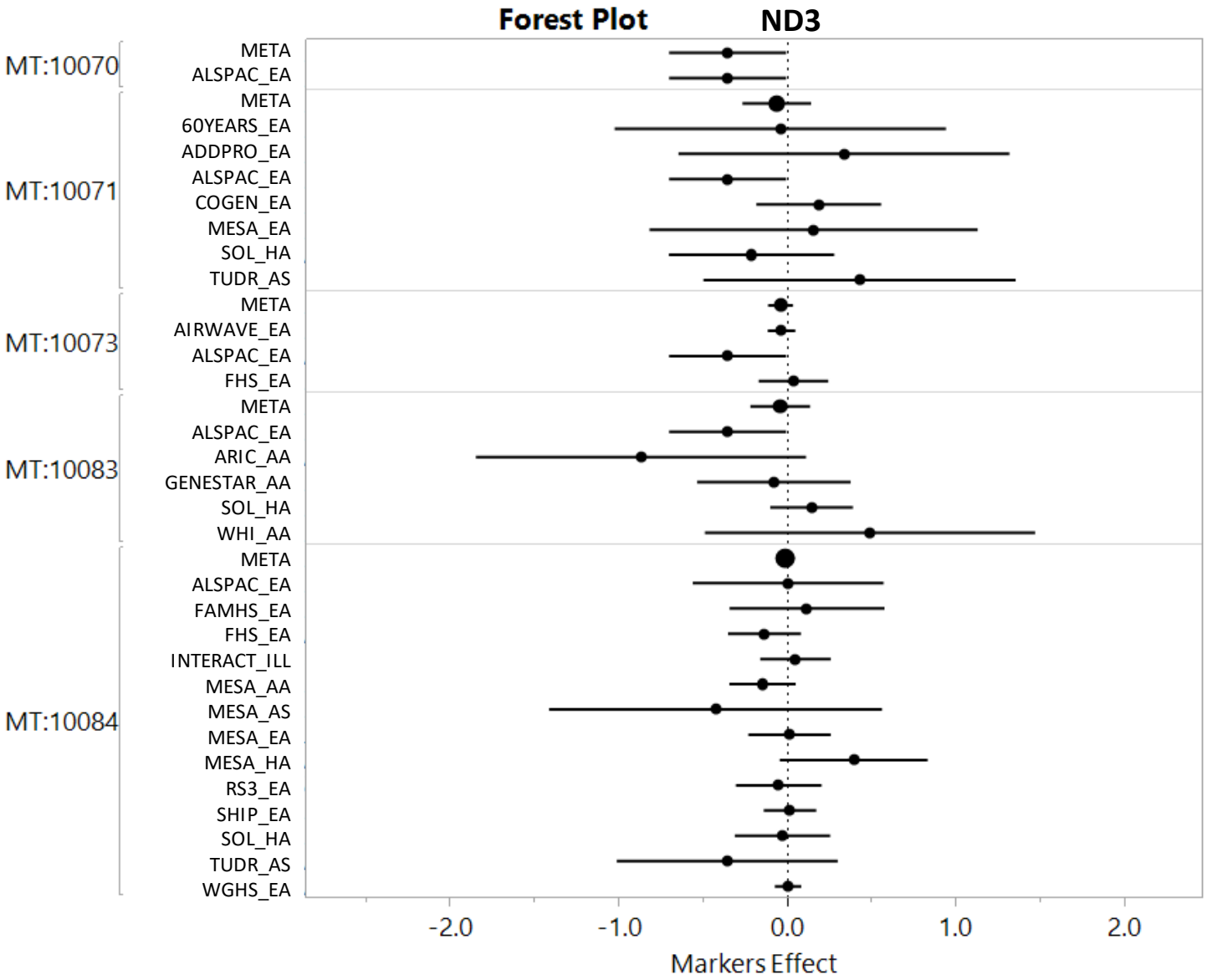


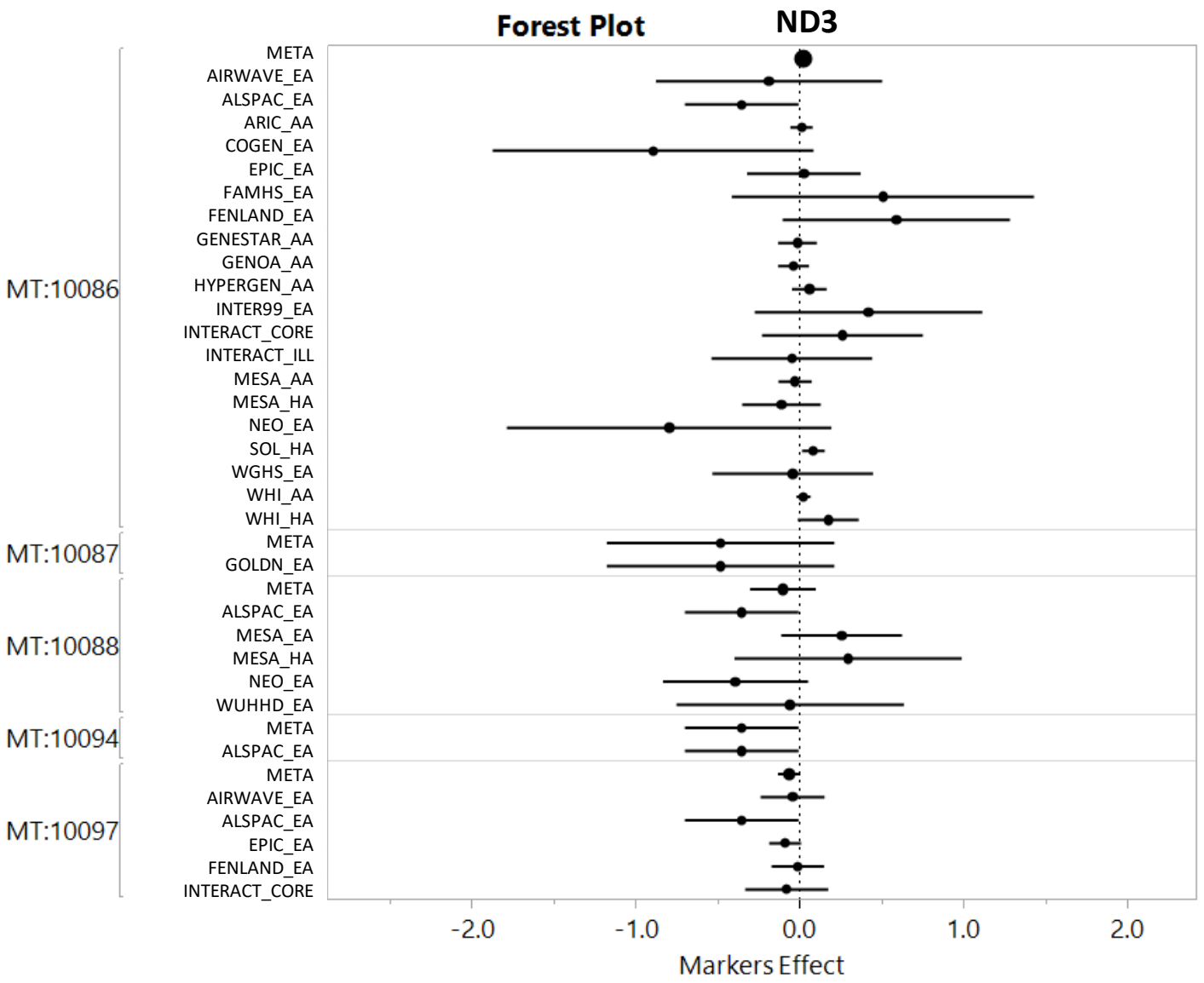
Figure S16. Forest plot of beta coefficients and their 95% confidence interval for studies contributing to seven significant mtDNA variants

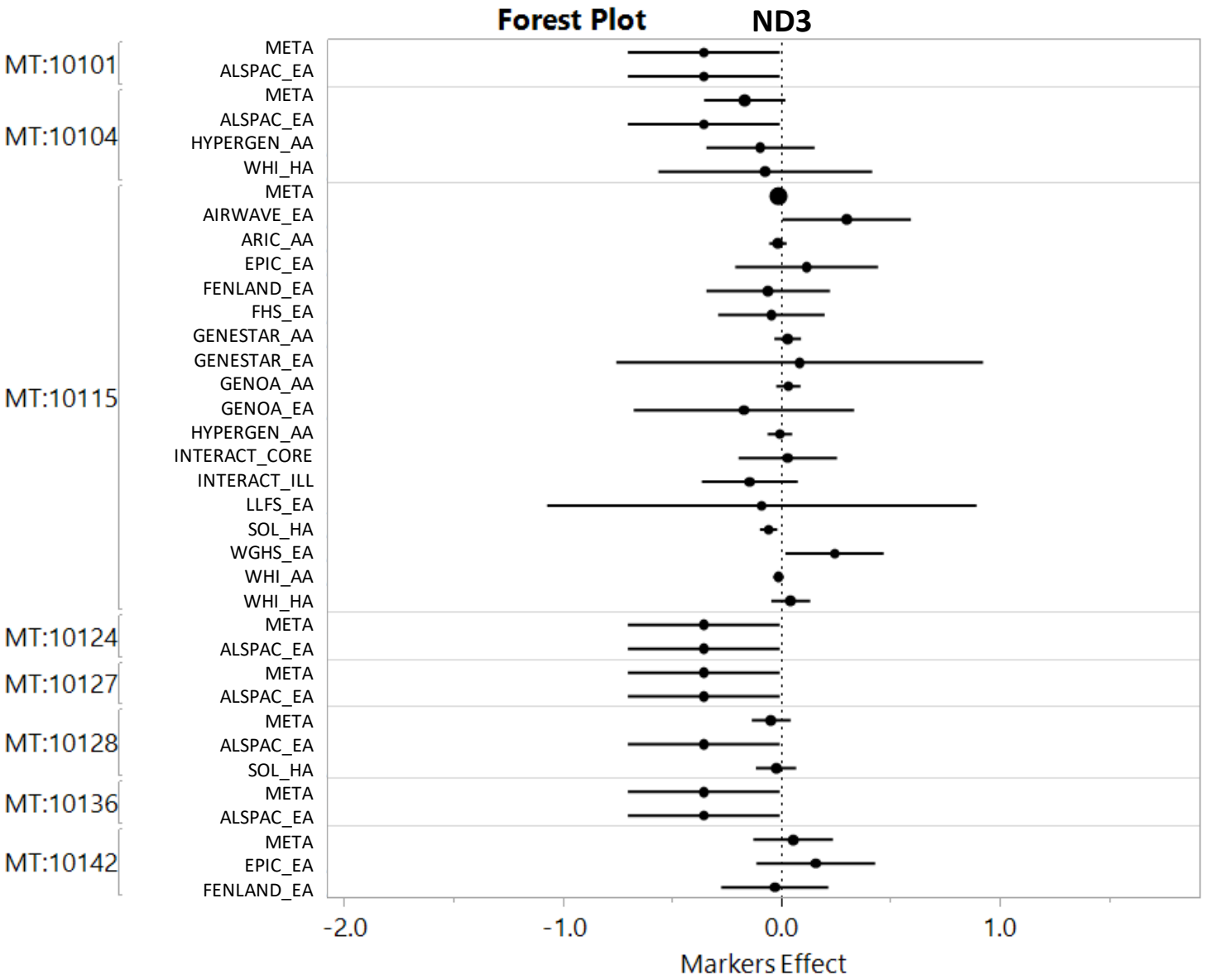


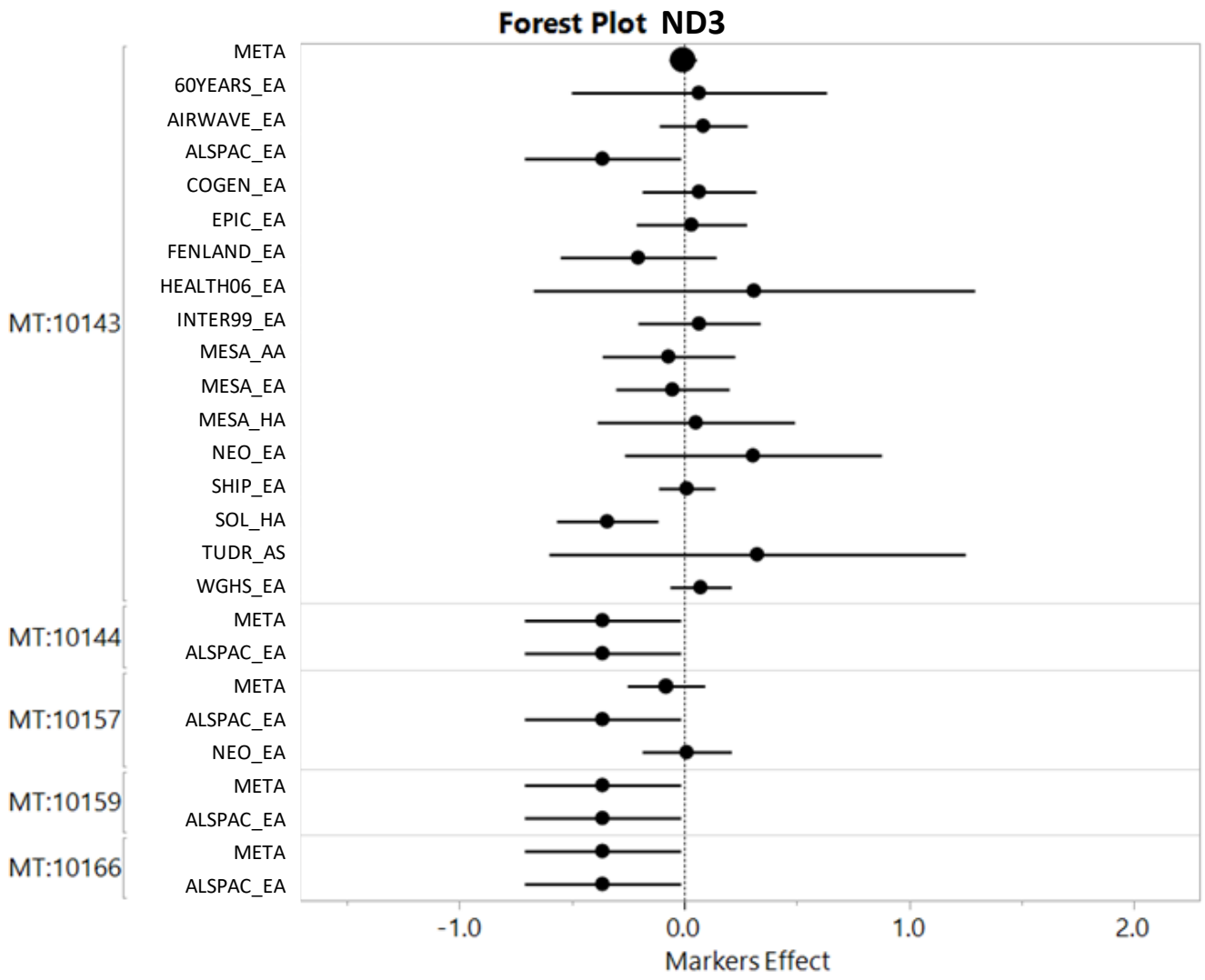
Variants beta-effects and their 95% CIs

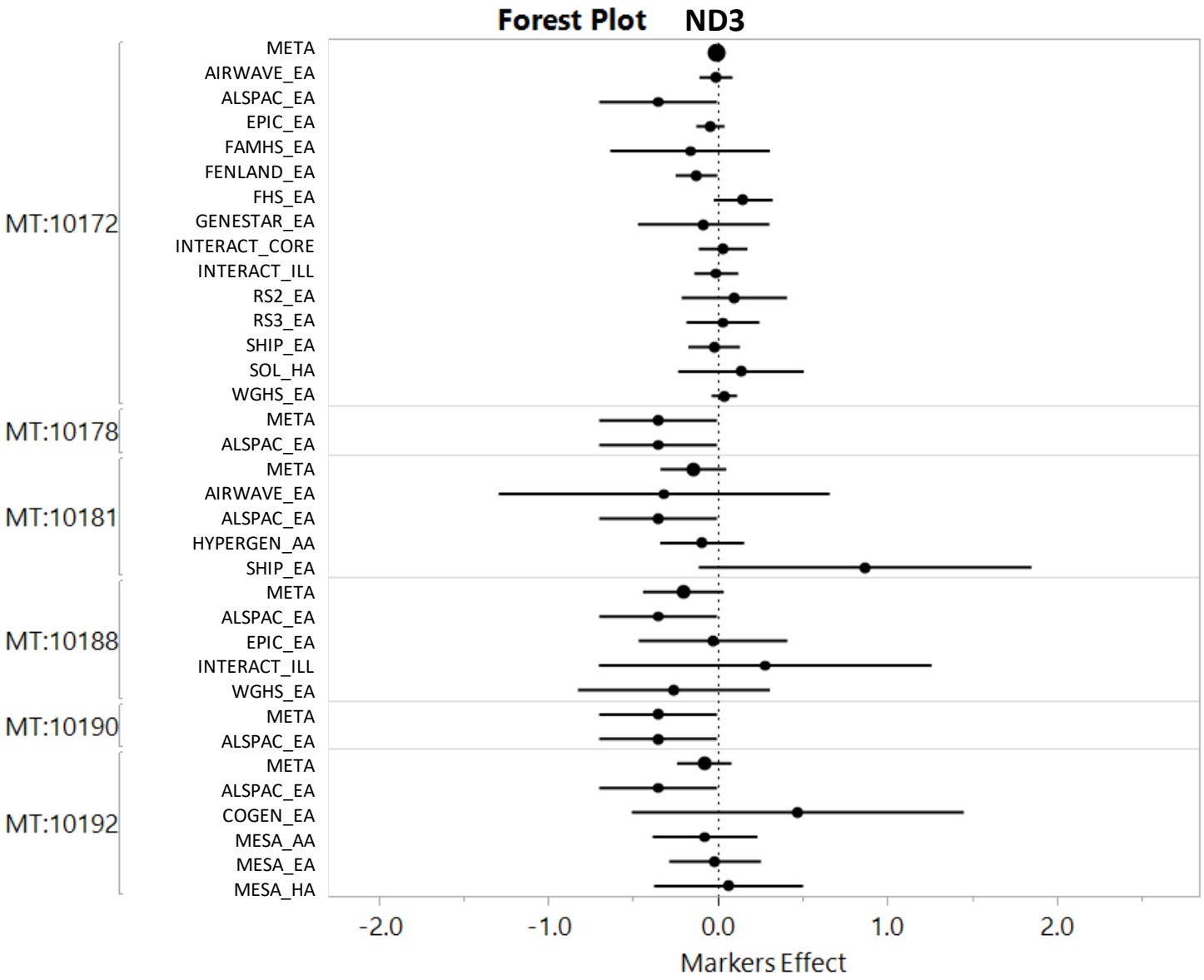
Figure S17. Forest plot of beta coefficients and their 95% confidence interval for studies contributing to MT-ND3-Gene based meta-analysis of 82 rare SNVs

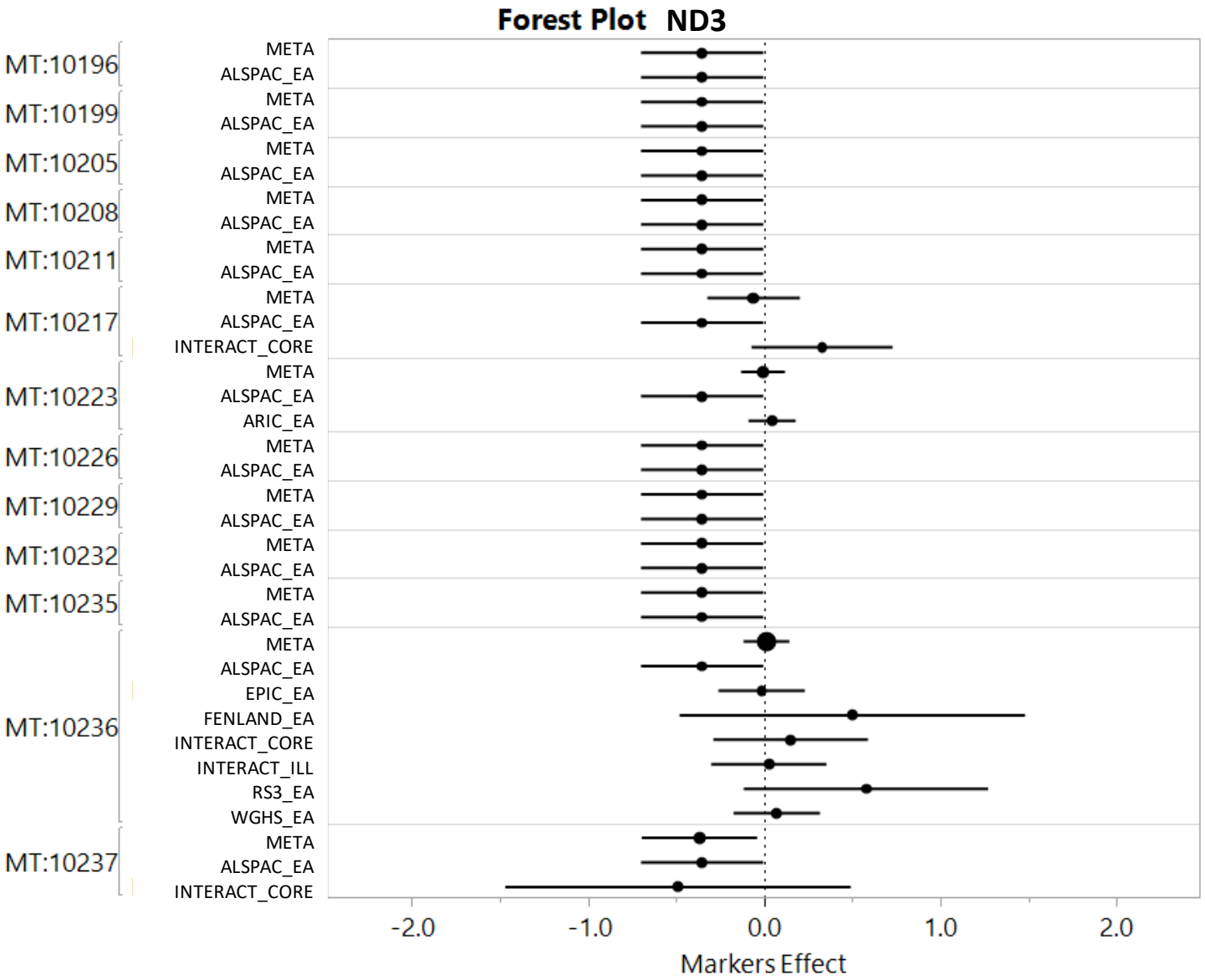

Forest Plot
ND3
MT:10070
META
ALSPAC_EA
MT:10071
META
60YEARS_EA
ADDPRO_EA
ALSPAC_EA
COGEN_EA
MESA_EA
SOL_HA
TUDR_AS
MT:10073
META
AIRWAVE_EA
ALSPAC_EA
FHS_EA
MT:10083
META
ALSPAC_EA
ARIC_AA
GENESTAR_AA
SOL_HA
WHI_AA
MT:10084
META
ALSPAC_EA
FAMHS_EA
FHS_EA
INTERACT_ILL
MESA_AA
MESA_AS
MESA_EA
MESA_HA
RS3_EA
SHIP_EA
SOL_HA
TUDR_AS
WGHS_EA
-2.0
-1.0
0.0
1.0
2.0
Markers Effect

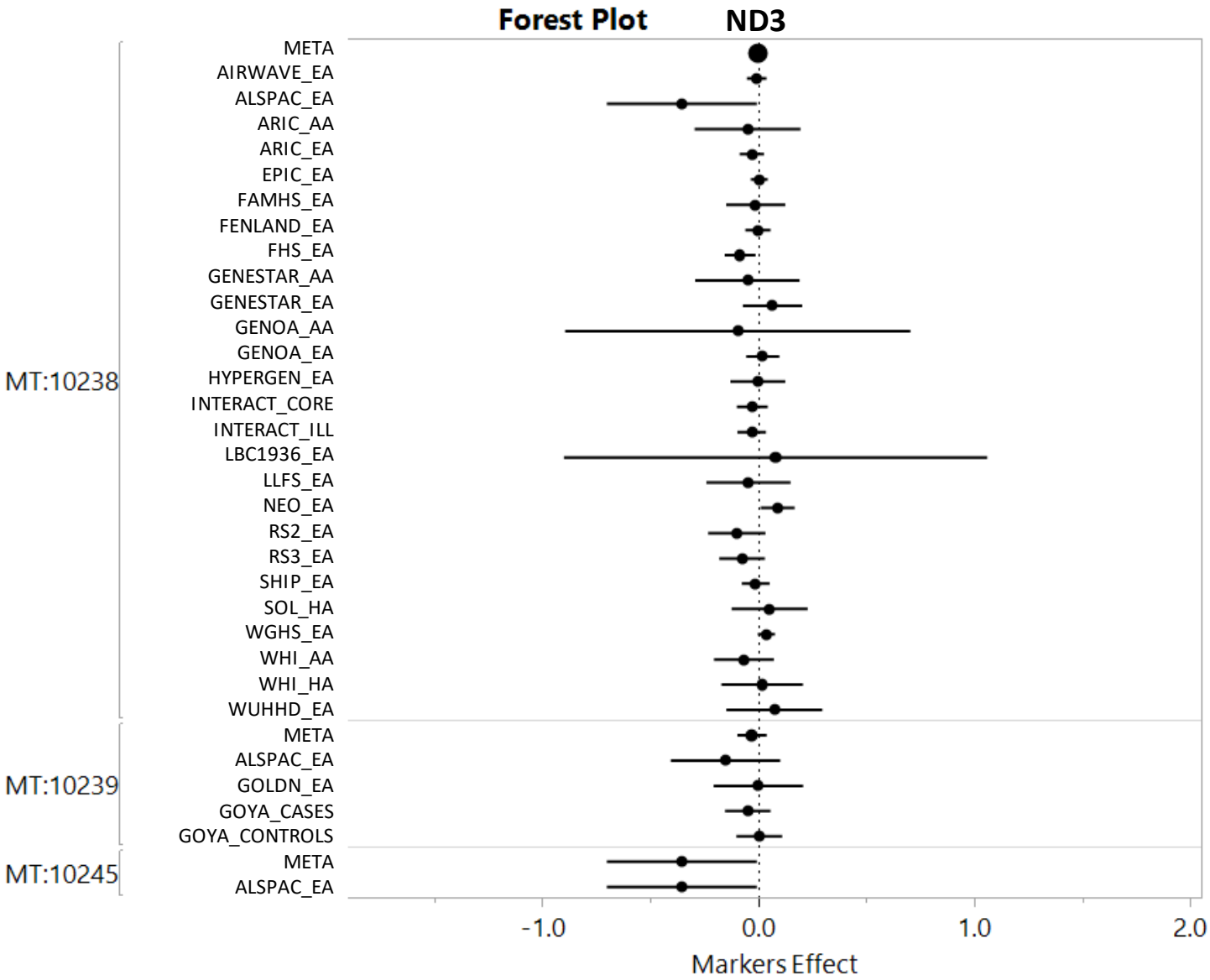

Forest Plot
ND3
MT:10086
META
AIRWAVE_EA
ALSPAC_EA
ARIC_AA
COGEN_EA
EPIC_EA
FAMHS_EA
FENLAND_EA
GENESTAR_AA
GENOA_AA
HYPERGEN_AA
INTER99_EA
INTERACT_CORE
INTERACT_ILL
MESA_AA
MESA_HA
NEO_EA
SOL_HA
WGHS_EA
WHI_AA
WHI_HA
MT:10087
META
GOLDN_EA
MT:10088
META
ALSPAC_EA
MESA_EA
MESA_HA
NEO_EA
WUHHD_EA
MT:10094
META
ALSPAC_EA
MT:10097
META
AIRWAVE_EA
ALSPAC_EA
EPIC_EA
FENLAND_EA
INTERACT_CORE
-2.0
-1.0
0.0
1.0
2.0
Markers Effect

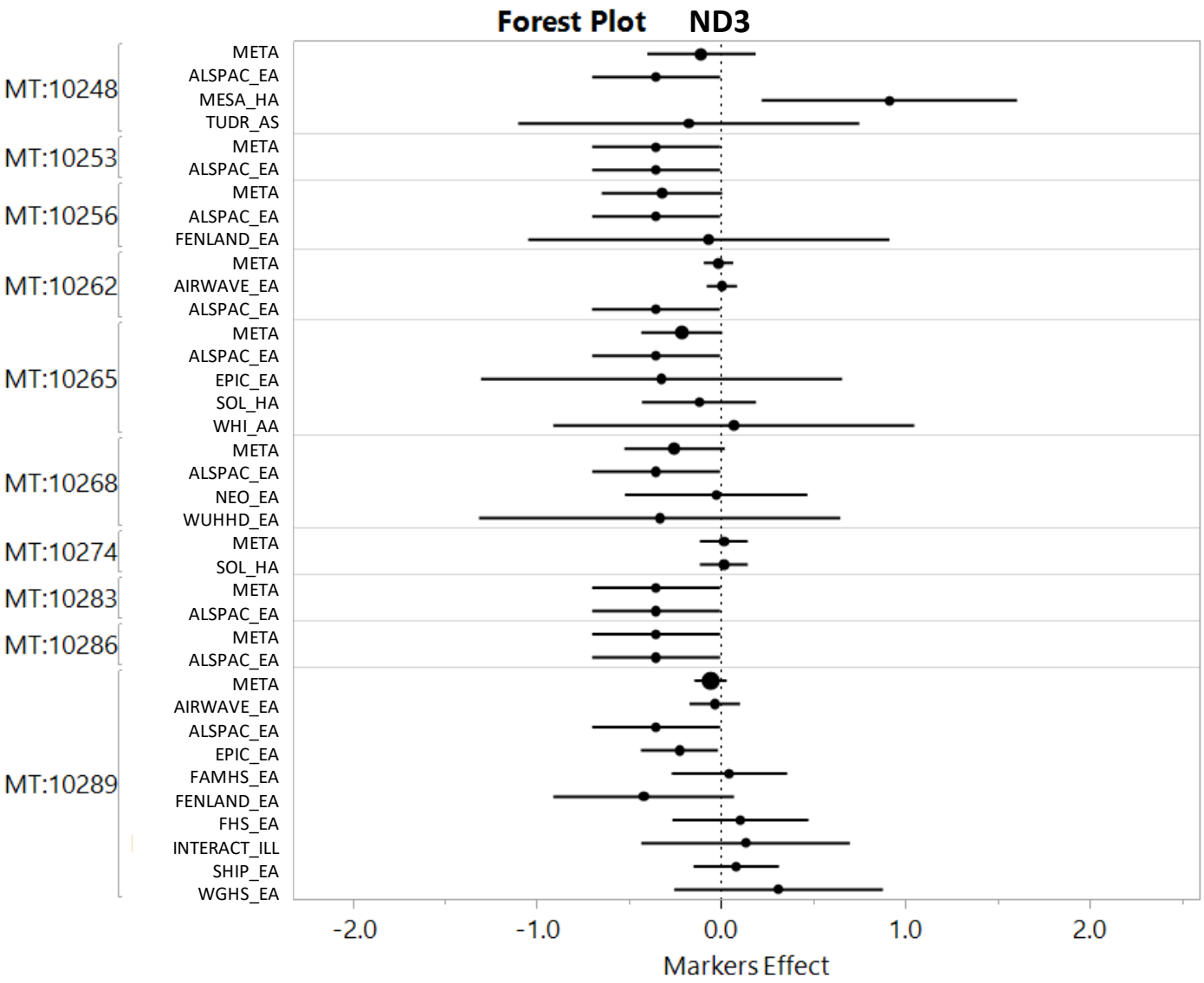

Forest Plot
ND3
MT:10101
META
ALSPAC_EA
MT:10104
META
ALSPAC_EA
HYPERGEN_AA
WHI_HA
MT:10115
META
AIRWAVE_EA
ARIC_AA
EPIC_EA
FENLAND_EA
FHS_EA
GENESTAR_AA
GENESTAR_EA
GENOA_AA
GENOA_EA
HYPERGEN_AA
INTERACT_CORE
INTERACT_ILL
LLFS_EA
SOL_HA
WGHS_EA
WHI_AA
WHI_HA
MT:10124
META
ALSPAC_EA
MT:10127
META
ALSPAC_EA
MT:10128
META
ALSPAC_EA
SOL_HA
MT:10136
META
ALSPAC_EA
MT:10142
META
EPIC_EA
FENLAND_EA
-2.0
-1.0
0.0
1.0
Markers Effect

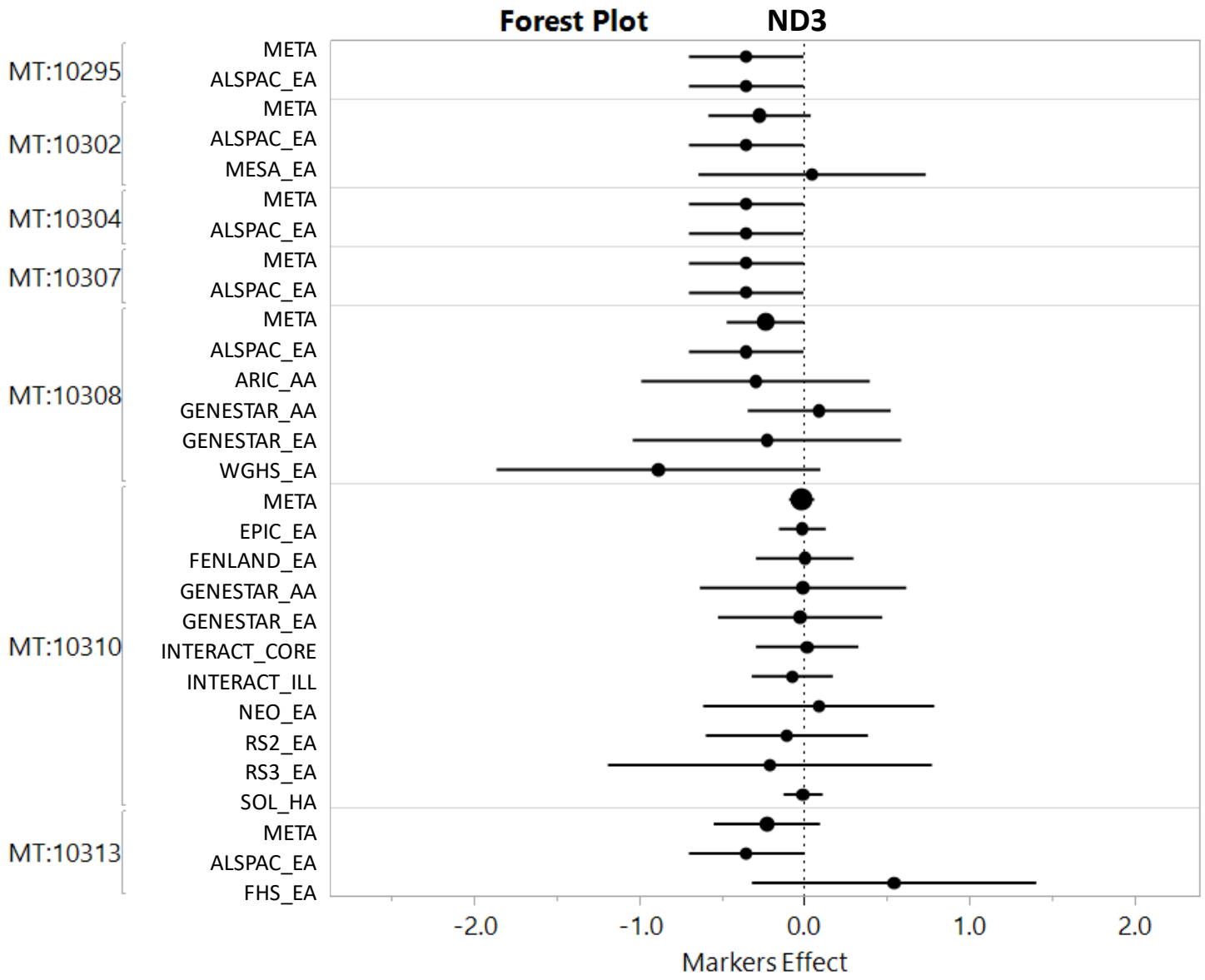

Forest Plot ND3
META
60YEARS_EA
AIRWAVE_EA
ALSPAC_EA
COGEN_EA
EPIC_EA
FENLAND_EA
HEALTH06_EA
MT:10143
INTER99_EA
MESA_AA
MESA_EA
MESA_HA
NEO_EA
SHIP_EA
SOL_HA
TUDR_AS
WGHS_EA
META
MT:10144
ALSPAC_EA
META
MT:10157
ALSPAC_EA
NEO_EA
META
MT:10159
ALSPAC_EA
META
MT:10166
ALSPAC_EA
-1.0
0.0
1.0
2.0
Markers Effect

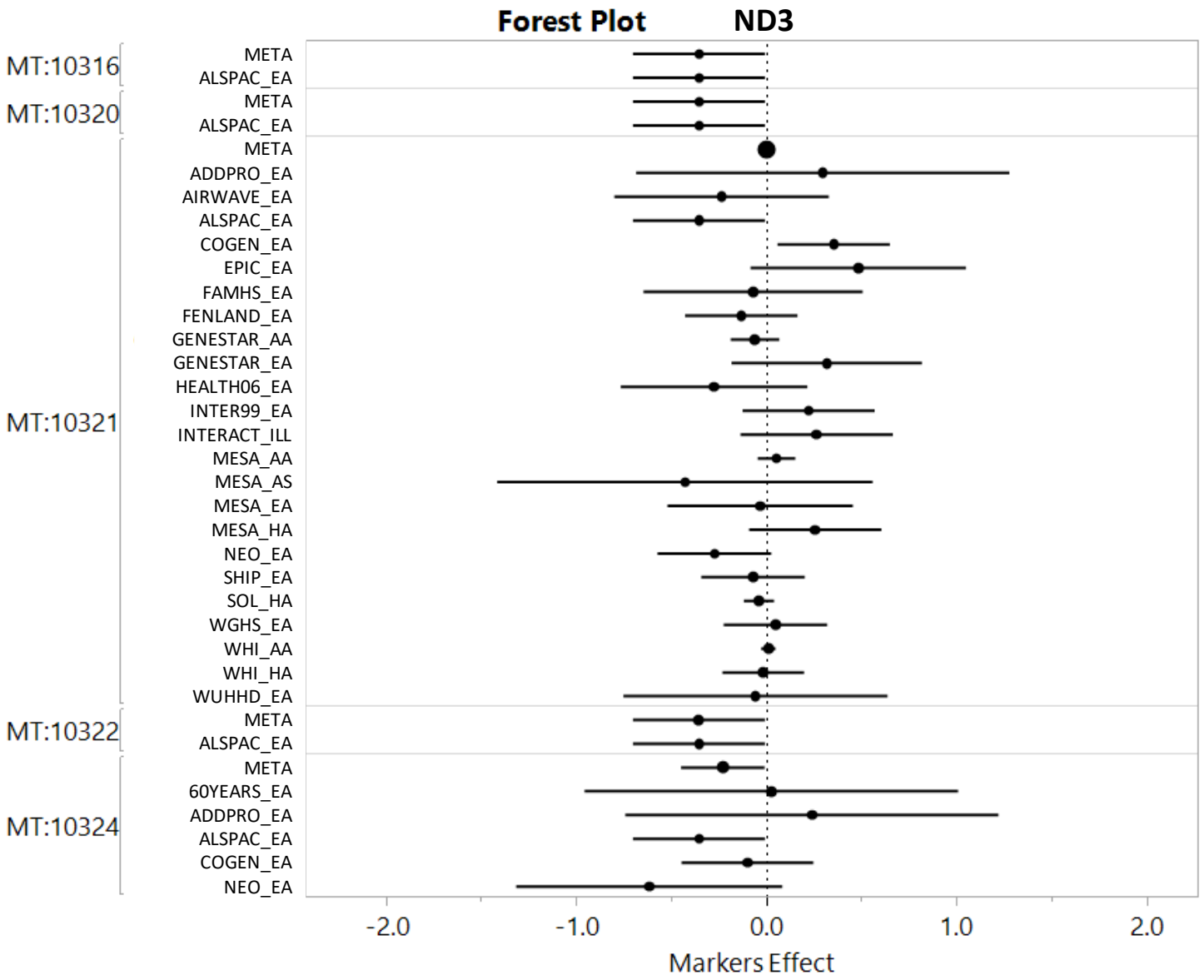

Forest Plot ND3
MT:10172
META
AIRWAVE_EA
ALSPAC_EA
EPIC_EA
FAMHS_EA
FENLAND_EA
FHS_EA
GENESTAR_EA
INTERACT_CORE
INTERACT_ILL
RS2_EA
RS3_EA
SHIP_EA
SOL_HA
WGHS_EA
MT:10178
META
ALSPAC_EA
MT:10181
META
AIRWAVE_EA
ALSPAC_EA
HYPERGEN_AA
SHIP_EA
MT:10188
META
ALSPAC_EA
EPIC_EA
INTERACT_ILL
WGHS_EA
MT:10190
META
ALSPAC_EA
MT:10192
META
ALSPAC_EA
COGEN_EA
MESA_AA
MESA_EA
MESA_HA
-2.0
-1.0
0.0
1.0
2.0
Markers Effect

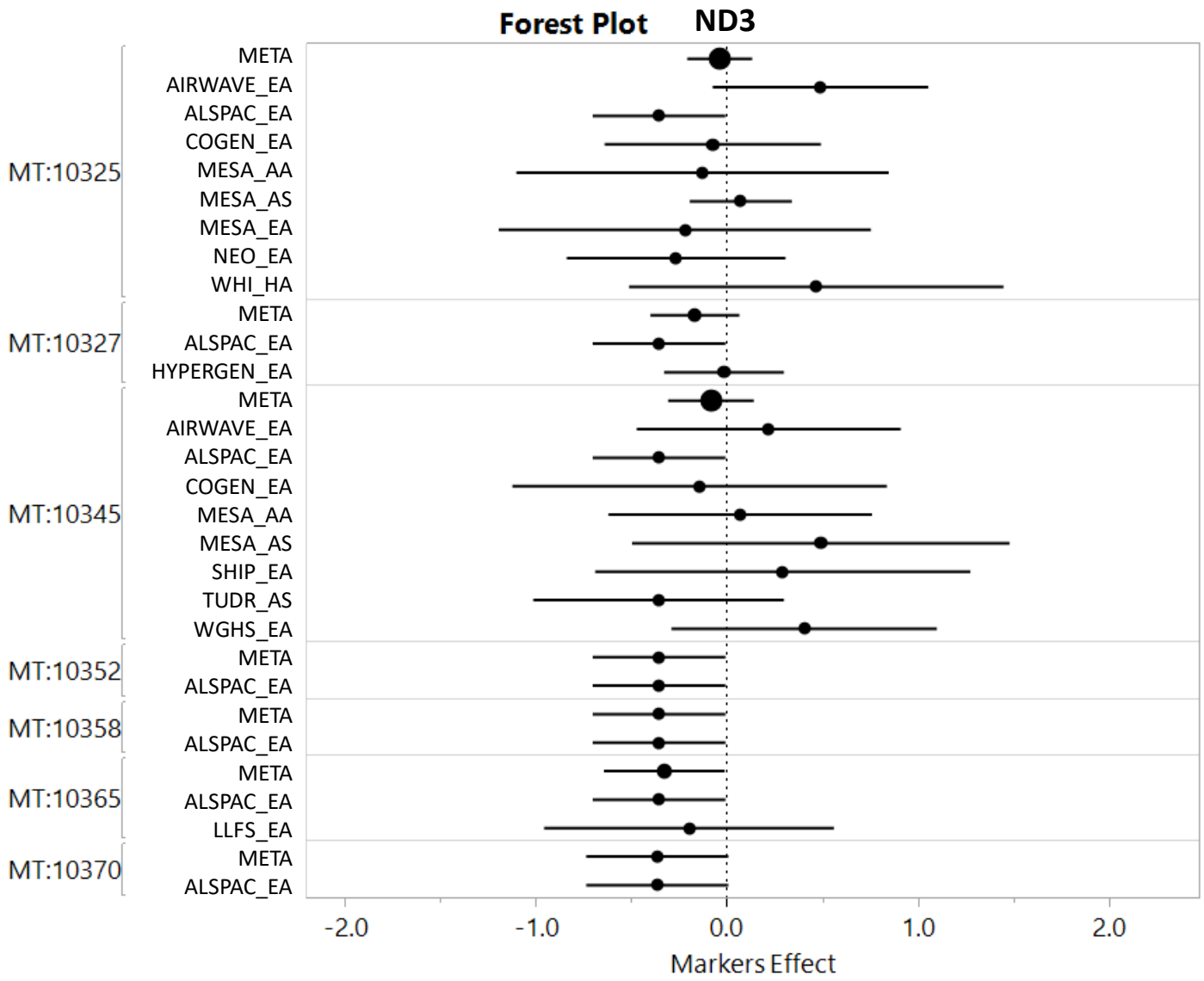

Forest Plot ND3
MT:10196
META
ALSPAC_EA
MT:10199
META
ALSPAC_EA
MT:10205
META
ALSPAC_EA
MT:10208
META
ALSPAC_EA
MT:10211
META
ALSPAC_EA
MT:10217
META
ALSPAC_EA
INTERACT_CORE
MT:10223
META
ALSPAC_EA
ARIC_EA
MT:10226
META
ALSPAC_EA
MT:10229
META
ALSPAC_EA
MT:10232
META
ALSPAC_EA
MT:10235
META
ALSPAC_EA
MT:10236
META
ALSPAC_EA
EPIC_EA
FENLAND_EA
INTERACT_CORE
INTERACT_ILL
RS3_EA
WGHS_EA
MT:10237
META
ALSPAC_EA
INTERACT_CORE
-2.0
-1.0
0.0
1.0
2.0
Markers Effect

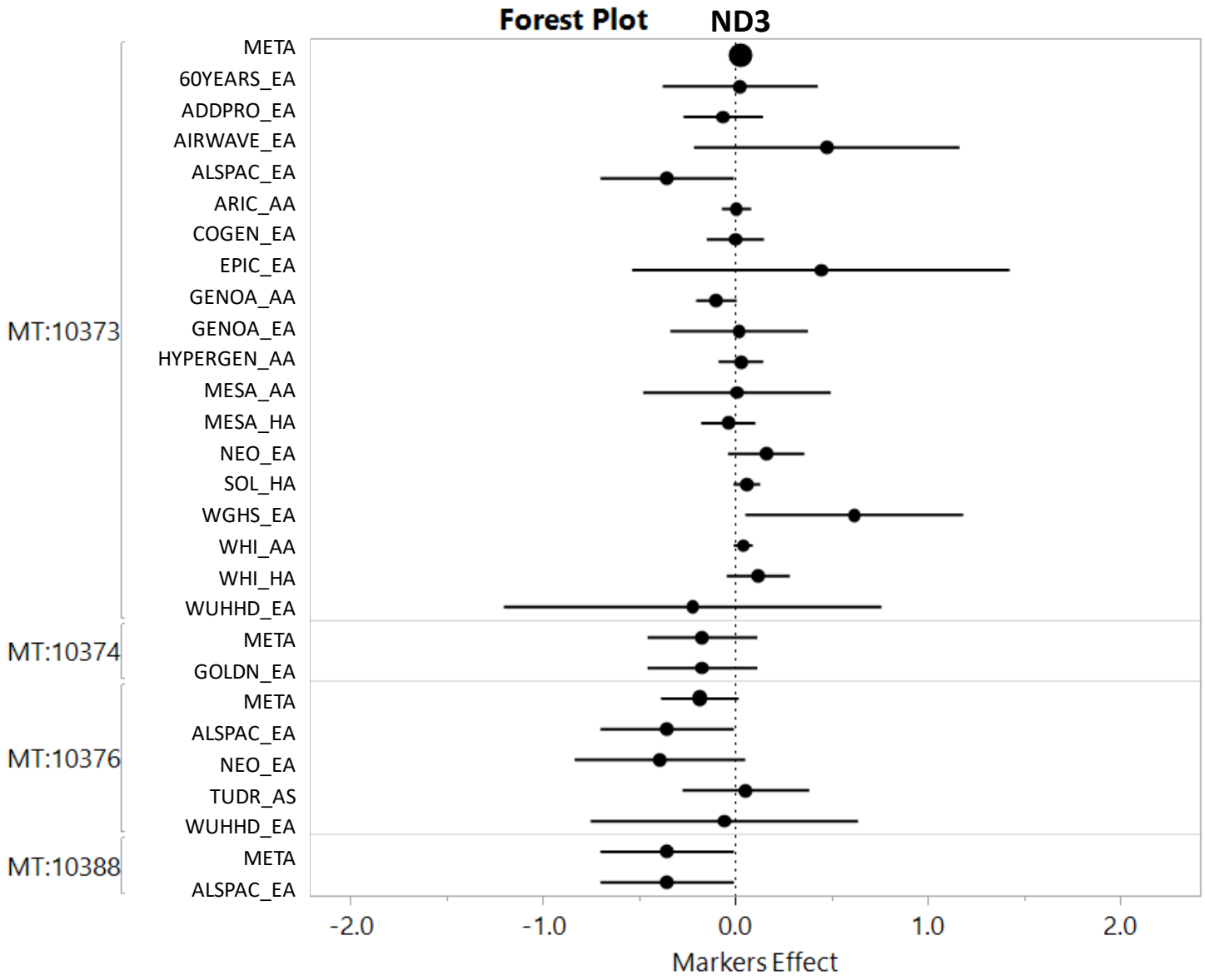

Forest Plot
ND3
META
AIRWAVE_EA
ALSPAC_EA
ARIC_AA
ARIC_EA
EPIC_EA
FAMHS_EA
FENLAND_EA
FHS_EA
GENESTAR_AA
GENESTAR_EA
GENOA_AA
GENOA_EA
HYPERGEN_EA
INTERACT_CORE
INTERACT_ILL
LBC1936_EA
LLFS_EA
NEO_EA
RS2_EA
RS3_EA
SHIP_EA
SOL_HA
WGHS_EA
WHI_AA
WHI_HA
WUHHD_EA
META
ALSPAC_EA
GOLDN_EA
GOYA_CASES
GOYA_CONTROLS
META
ALSPAC_EA
MT:10238
MT:10239
MT:10245
-1.0
0.0
1.0
2.0
Markers Effect

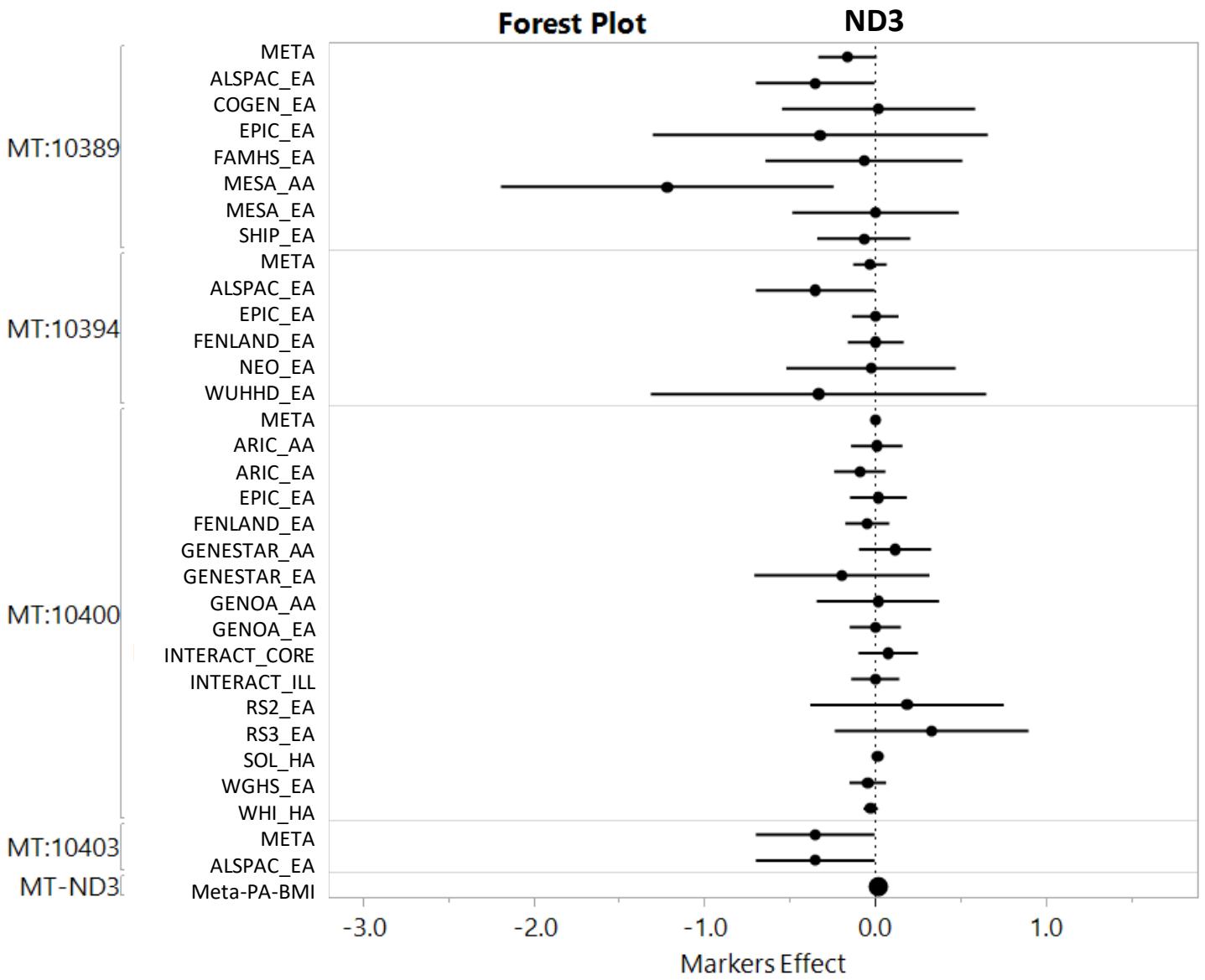

Forest Plot ND3
MT:10248
META
ALSPAC_EA
MESA_HA
TUDR_AS
MT:10253
META
ALSPAC_EA
MT:10256
META
ALSPAC_EA
FENLAND_EA
MT:10262
META
AIRWAVE_EA
ALSPAC_EA
MT:10265
META
ALSPAC_EA
EPIC_EA
SOL_HA
WHI_AA
MT:10268
META
ALSPAC_EA
NEO_EA
WUHHD_EA
MT:10274
META
SOL_HA
MT:10283
META
ALSPAC_EA
MT:10286
META
ALSPAC_EA
MT:10289
META
AIRWAVE_EA
ALSPAC_EA
EPIC_EA
FAMHS_EA
FENLAND_EA
FHS_EA
INTERACT_ILL
SHIP_EA
WGHS_EA
-2.0
-1.0
0.0
1.0
2.0
Markers Effect


Forest Plot
ND3
MT:10295
META
ALSPAC_EA
MT:10302
META
ALSPAC_EA
MESA_EA
MT:10304
META
ALSPAC_EA
MT:10307
META
ALSPAC_EA
MT:10308
META
ALSPAC_EA
ARIC_AA
GENESTAR_AA
GENESTAR_EA
WGHS_EA
MT:10310
META
EPIC_EA
FENLAND_EA
GENESTAR_AA
GENESTAR_EA
INTERACT_CORE
INTERACT_ILL
NEO_EA
RS2_EA
RS3_EA
SOL_HA
MT:10313
META
ALSPAC_EA
FHS_EA
-2.0
-1.0
0.0
1.0
2.0
Markers Effect


Forest Plot
ND3
MT:10316
META
ALSPAC_EA
MT:10320
META
ALSPAC_EA
MT:10321
META
ADDPRO_EA
AIRWAVE_EA
ALSPAC_EA
COGEN_EA
EPIC_EA
FAMHS_EA
FENLAND_EA
GENESTAR_AA
GENESTAR_EA
HEALTH06_EA
INTER99_EA
INTERACT_ILL
MESA_AA
MESA_AS
MESA_EA
MESA_HA
NEO_EA
SHIP_EA
SOL_HA
WGHS_EA
WHI_AA
WHI_HA
WUHHD_EA
MT:10322
META
ALSPAC_EA
MT:10324
META
60YEARS_EA
ADDPRO_EA
ALSPAC_EA
COGEN_EA
NEO_EA
-2.0
-1.0
0.0
1.0
2.0
Markers Effect


Forest Plot
ND3
MT:10325
META
AIRWAVE_EA
ALSPAC_EA
COGEN_EA
MESA_AA
MESA_AS
MESA_EA
NEO_EA
WHI_HA
MT:10327
META
ALSPAC_EA
HYPERGEN_EA
MT:10345
META
AIRWAVE_EA
ALSPAC_EA
COGEN_EA
MESA_AA
MESA_AS
SHIP_EA
TUDR_AS
WGHS_EA
MT:10352
META
ALSPAC_EA
MT:10358
META
ALSPAC_EA
MT:10365
META
ALSPAC_EA
LLFS_EA
MT:10370
META
ALSPAC_EA
-2.0
-1.0
0.0
1.0
2.0
Markers Effect


Forest Plot
ND3
META
60YEARS_EA
ADDPRO_EA
AIRWAVE_EA
ALSPAC_EA
ARIC_AA
COGEN_EA
EPIC_EA
GENOA_AA
GENOA_EA
HYPERGEN_AA
MESA_AA
MESA_HA
NEO_EA
SOL_HA
WGHS_EA
WHI_AA
WHI_HA
WUHHD_EA
MT:10373
META
GOLDN_EA
MT:10374
META
ALSPAC_EA
NEO_EA
TUDR_AS
WUHHD_EA
MT:10376
META
ALSPAC_EA
MT:10388
-2.0
-1.0
0.0
1.0
2.0
Markers Effect


Forest Plot
ND3
MT:10389
META
ALSPAC_EA
COGEN_EA
EPIC_EA
FAMHS_EA
MESA_AA
MESA_EA
SHIP_EA
MT:10394
META
ALSPAC_EA
EPIC_EA
FENLAND_EA
NEO_EA
WUHHD_EA
MT:10400
META
ARIC_AA
ARIC_EA
EPIC_EA
FENLAND_EA
GENESTAR_AA
GENESTAR_EA
GENOA_AA
GENOA_EA
INTERACT_CORE
INTERACT_ILL
RS2_EA
RS3_EA
SOL_HA
WGHS_EA
WHI_HA
MT:10403
META
ALSPAC_EA
MT-ND3
Meta-PA-BMI
-3.0
-2.0
-1.0
0.0
1.0
Markers Effect

2. **Supplementary Tables**

Table S1. Summary Statistics for 7 Phenotypes Studied by Cohort

Ancestry	Cohort	BMI			WHR			GLUC			INS			HOMAB			HOMAIR			HBA1C		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
European	AIRWAVE	16025	27.2	4.1	15979	0.86	0.07	958	95.5	9.5										16012	5.58	0.39
	ALSPAC	3628	22.7	3.9	5303	0.84	0.06	2761	93.76	6.39	2767	10.01	4.92	2752	108.53	55.65	2752	2.33	1.21	1456	4.91	0.29
	ARIC	9481	27	4.86	9473	0.93	0.08	8651	98.73	9.21	9277	13.45	13.25	8646	112.91	74.94	8648	3.05	2.47			
	DANISH-60YEARS	642	26.58	3.76	651	0.90	0.09	612	94.42	9.72	612	7.18	4.07	612	76.54	38.17	612	1.72	1.11			
	DANISH-ADDPRO	1505	27.17	4.18	1532	0.94	0.08	1264	105.68	9.62	1262	7.16	4.87	1262	55.78	41.05	1262	1.90	1.38	1539	5.75	0.47
	DANISH-COGEN	5653	27.26	4.40				1843	103.18	12.38										557	6.16	1.80
	DANISH-GOYA_CASES	3197	35.54	2.90	787	1.02	0.07	197	103.80	9.01	196	11.06	7.71	196	91.57	62.02	196	2.88	2.09	224	5.94	1.02
	DANISH-GOYA_CONTROLS	3243	22.34	2.71	910	0.93	0.06	307	101.15	8.57	302	6.39	4.68	302	55.39	37.50	302	1.63	1.26	317	5.63	0.43
	DANISH-HEALTH06	2815	25.44	3.89	2907	0.87	0.09	2616	97.55	9.73	2616	6.81	4.65	2616	64.33	95.72	2616	1.68	1.25	2906	5.43	0.48
	DANISH-INTER99	5979	25.86	3.86	6136	0.86	0.09	5723	98.02	9.20	5506	6.85	4.50	5500	66.08	59.27	5500	1.68	1.16	6153	5.86	0.63
	DANISH-VEJLE	2367	28.65	5.41	2392	0.95	0.08													2408	6.68	1.13
	EPIC-INTERACT_CORE	6745	27.26	4.79	6036	0.87	0.1	632	90.33	12.71										4056	5.42	0.35
	EPIC-INTERACT_ILL	8364	27.92	4.96	7583	0.88	0.09	714	90.07	13.62										4065	5.43	0.34
	EPIC-NORFOLK	19225	26.29	3.82	19199	0.86	0.09													8499	5.19	0.55
	FAMHS	3677	27.74	5.47	3674	0.92	0.09	3362	94.04	9.86	3610	11.57	13.43	3353	116.10	211.22	3356	2.48	2.10			
	FENLAND	8453	26.91	4.83	8437	0.88	0.09	8407	86.43	8.75	7084	7.96	6.51	7066	112.68	85.9	7066	1.74	1.57	5377	5.52	0.33
	FHS	6930	27.44	5.46	6046	0.93	0.08	6423	93.51	8.7	5984	29.85	16.5	5982	364.97	218.5	5984	7.07	4.31	5175	5.46	0.3
	GENESTAR	1887	28.37	5.88	337	0.89	0.097	1577	89.85	10.05	651	9	7.41	646	123.38	102.7	650	2.03	1.87			
	GENOA	1016	30.77	6.20	1012	0.91	0.10	1042	90.69	7.91	1021	8.01	5.09	1021	94.97	55.70	1021	1.67	1.13			
	GOLDN	842	28.47	5.49	842	0.90	0.10	842	102.01	19.53	841	14.07	8.23	841	126.88	63.01	835	3.61	2.47			
	HYPERGEN	1267	29.42	6.14	1251	0.91	0.09	1105	94.25	9.79	1103	7.49	5.3	985	86.74	33.9	985	1.06	0.68			
	LBC1921	513	26.19	4.09																452	5.7	0.7
	LBC1936	1004	27.79	4.32																999	5.93	0.74
	LLFS	4401	27.13	4.85				3840	91.73	11.42	3812	7.93	5.51	3779	97.33	79.86	3799	1.85	1.44	4163	5.53	0.35
	MESA	2378	27.52	4.93	2378	0.92	0.09	2372	87.89	10.14	2372	8.76	4.98	2370	120.36	69.31	2370	1.95	1.26	2217	5.34	0.37
	NEO	5744	29.99	4.83	5739	0.92	0.08	5096	98.53	9.71	5089	11.75	7.56	5089	110.97	69.82	5089	2.92	2.03	5097	5.34	0.26
	OOA	2296	26.73	5.03	2231	0.87	0.07	2023	85.76	9.47	829	10.45	5.33	752	153.43	113.8	759	2.35	1.41	720	5.14	0.54
	PELOTAS	850	26.87	5.34	852	0.81	0.08													842	5.06	0.41
	RS-II	2152	27.23	3.99	1938	0.91	0.09	1758	100.6	9.14	1744	12.74	6.8	1744	114.57	58.89	1744	3.22	1.85			
	RS-III	3026	27.7	4.59	2926	0.87	0.08	2740	95.16	10.09	2665	14.94	9.38	2643	157.71	119.1	2658	3.59	2.49			
	SHIP	8205	27.73	5.01	8196	0.88	0.09	2741	95.98	10.47	477	1.58	0.58	471	3.9	0.56	477	0.17	0.62	6982	5.19	0.51
	WGHS	22203	25.88	4.88	19568	0.83	0.08													21970	5.01	0.28
	WUHHD	710	29.82	6.38	681	0.87	0.08	558	87.34	12.48	567	9.4	7.87	541	142.05	160.3	548	2.1	1.88			
African	ARIC	2860	29.7	6.06	2857	0.92	0.08	2250	98.73	10.13	2587	19.32	28.42	2244	152.5	103.1	2249	4.1	3.49			
	GENESTAR	1165	31.36	7.52	552	0.88	0.076	949	90.49	11.79	500	10.78	7.04	496	141.03	109.2	499	2.49	1.83			
	GENOA	1003	31.74	6.57	1001	0.90	0.06	688	91.56	8.28	687	9.22	7.95	687	106.80	94.22	687	1.94	1.72			
	HYPERGEN	1256	32.54	8.02	1242	0.9	0.08	991	93.01	10.49	989	10.03	8.03	904	105.25	42.38	904	1.37	0.88			
	MESA	1345	29.74	5.77	1345	0.91	0.08	1339	90.13	10.83	1336	9.84	5.92	1335	122.97	97.94	1335	2.24	1.51	1168	5.56	0.44
	WHI	8116	31.03	6.50	8088	0.82	0.07	5514	91.84	9.98	5364	8.94	5.64	5357	104.34	66.04	5366	2.08	1.46			
Asian	MESA	672	23.83	3.2	672	0.91	0.07	670	91.47	9.78	671	8.91	4.61	669	108.54	64.67	669	2.03	1.13	605	5.51	0.37
	TUDR	895	24.75	4.33	168	0.93	0.05													835	8.88	2.47
Hispanic	HCHS-SOL	10120	29.73	5.71	9013	0.92	0.07	8108	94.64	8.79	8071	12.23	8.50	8077	128.86	86.01	8073	2.91	2.16	8106	5.46	0.38
	MESA	1187	29.04	4.88	1187	0.95	0.07	1185	90.78	10.86	1185	11.03	14.24	1184	135.77	129.6	1184	2.54	3.89	1063	5.49	0.41
	WHI	3463	28.76	5.36	3448	0.82	0.07	2623	91.55	8.88	2572	8.22	5.45	2568	95.55	60.25	2572	1.90	1.39			
Brazilian	PELOTAS	1925	26.92	5.59	1943	0.82	0.08													1921	5.09	0.43

Table S2(a,b). Details by Cohort of mtDNA Variants Associated with BMI, WHR, Glucose, Insulin, HOMA-B, HOMA-IR and HbA1c METAL Meta-Analysis **Single Variant** Results

a. Meta-analysis results of variants with MAF > 1% (see Table 1.a) including information at the cohort level selected with MAC ≥ 5.

No	Results	Pos	rsID	Gene	Annotation	Trait	Ancestry	A1/2	Freq1	MAF	MAC	INFO	β(SE)	*P*-value	Dir	Het-*P*	N	Missing Rate
1	Metal	8706		MT-ATP6		WHR	PA	A/G	0.9676	0.0324	834	n.a.	-0.13(0.04)	**4.07E-04**	---+	9.44E-02	25,748	n.a.
	ALSPAC.EA								0.9974	0.0026	14	0.816	-0.40(0.26)	1.27E-01			5,303	0.00000
	SOL.HA								0.9880	0.0120	108	0.960	-0.32(0.10)	8.05E-04			8,989	0.00266
	WHI.AA								0.9135	0.0865	693	0.968	-0.09(0.04)	2.10E-02			8,011	0.00952
	WHI.HA								0.9945	0.0055	19	0.447	0.01(0.23)	9.68E-01			3,445	0.00087
2	Metal	16320	rs62581338	D-loop		GLUC	PA	T/C	0.0158	0.0158	301	n.a.	-0.21(0.05)	**7.59E-05**	--+--	1.77E-01	19,046	n.a.
	FENLAND.EA								0.0028	0.0028	23	0.432	-0.13(0.10)	2.19E-01			8,209	0.02300
	GOYA_CASES.EA								0.2439	0.2439	40	1.000	-0.36(0.18)	5.24E-02			164	0.00000
	GOYA_CONTROLS.EA								0.2402	0.2402	61	1.000	0.01(0.15)	9.63E-01			254	0.00000
	SOL.HA								0.0195	0.0195	153	0.791	-0.33(0.08)	4.63E-05			7,849	0.03194
	WHI.HA								0.0093	0.0093	24	0.704	-0.05(0.21)	8.19E-01			2,570	0.01908

Footnotes: No-order number; Pos - MT position in bps; Gene- gene name or region; rsID - rsID-name from NCBI dbSNP database when available; Annotation - role of the variants when available; Trait - one or more of seven traits studied; Ancestry – AA – African Americans, EA – European, HA- Hispanics or Latino, and PA - Pan-ancestry; A1/2 - the coded and non-coded alleles; Freq1 - allele frequency for coded allele; FreqSE - Standard error of allele frequency from METAL; MINFreq - a minimum allele frequency for contributing cohorts; MAXFreq - a maximum allele frequency for contributing cohorts; MAF - minor allele frequency; MAC - minor allele count, calculated as MAF*N; β(SE) - beta coefficient and the corresponding standard error; *P*-value - from single variant regression analysis; Dir - direction sign of contributing cohort's beta; Het-*P* - heterogeneity *P*-value test from METAL; N - individuals' sample contributing in a particular marker meta-analysis.

b. Meta-analysis results of variants with MAF < 1% (see Table 1.b) including information at the cohort level selected with MAC ≥ 5.

No	Results	Pos	rsID	Gene	Annotation	Trait	Ancestry	A1/2	Freq1	MAF	MAC	INFO	β(SE)	*P*-value	Dir	Het-*P*	N	Missing Rate
1	Metal	8896	rs202120082	MT-ATP6	missense	WHR	PA	A/G	0.0038	0.0038	134	n.a.	0.30(0.08)	**1.12E-04**	++++	8.80E-01	34,959	n.a.
	ALSPAC.EA								0.0026	0.0026	14	0.821	0.40(0.26)	1.27E-01			5,303	0.00000
	MESA.HA								0.0059	0.0059	6	1.000	0.28(0.23)	2.20E-01			1,075	0.00000
	SOL.HA								0.0121	0.0121	109	1.000	0.32(0.10)	7.66E-04			9,013	0.00000
	WGHS.EA								0.0003	0.0003	5	1.000	0.15(0.22)	5.09E-01			19,568	0.00000
2	Metal	14124		MT-ND5		INS	PA	T/C	0.0035	0.0035	21	n.a.	0.57(0.16)	**2.95E-04**	+++	1.62E-01	6,035	n.a.
	ALSPAC.EA								0.0022	0.0022	6	0.995	0.43(0.38)	2.51E-01			2,767	0.00037
	ARIC.AA								0.0035	0.0035	9	0.948	1.09(0.32)	5.80E-04			2,581	0.00070
	GENOA.AA								0.0087	0.0087	6	0.999	0.38(0.20)	5.97E-02			687	0.00000
3	Metal	14272	rs2853814	MT-ND6	missense	BMI	PA	T/C	0.0012	0.0012	17	n.a.	0.84(0.24)	**4.90E-04**	++	6.49E-01	13,636	n.a.
	ALSPAC.EA								0.0022	0.0022	8	0.994	0.73(0.35)	3.99E-02			3,628	0.00037
	SOL.HA								0.0009	0.0009	9	0.469	0.95(0.33)	4.38E-03			10,008	0.01107
4	Metal	14353		MT-ND6		WHR	PA	T/C	0.9916	0.0084	120	n.a.	-0.33(0.09)	**2.29E-04**	--	8.00E-01	14,315	n.a.
	ALSPAC.EA								0.9978	0.0022	12	0.994	-0.40(0.26)	1.27E-01			5,303	0.00037
	SOL.HA								0.9880	0.0120	108	0.967	-0.33(0.10)	7.75E-04			9,012	0.00011
5	Metal	14584		MT-ND6		BMI	PA	T/C	0.9988	0.0012	17	n.a.	-0.84(0.24)	**4.90E-04**	--	6.49E-01	13,636	n.a.
	ALSPAC.EA								0.9978	0.0022	8	0.994	-0.73(0.35)	3.99E-02			3,628	0.00037
	SOL.HA								0.9991	0.0009	9	0.469	-0.95(0.33)	4.38E-03			10,008	0.01107

Table S3. Association Results at Cohort Level for SOL-HA.

No	POS	rsID	Trait	Ancestry	A1	A2	Freq1	MAF	β	SE	*P*-value	N	missRate
1*	8706		WHR	Central American	A	G	0.924361	0.075639	-0.26	0.12	2.68E-02	1018	0.00196
				Cuban	A	G	0.992486	0.007514	-0.46	0.29	1.11E-01	1730	0.00518
				Dominican	NA	NA	NA	NA	NA	NA	NA	NA	NA
				Mexican	A	G	0.996458	0.003542	-0.46	0.29	1.13E-01	3106	0.00064
				Puerto Rican	NA	NA	NA	NA	NA	NA	NA	NA	NA
				South American	A	G	0.989766	0.010234	-0.93	0.37	1.24E-02	684	0.00437
2	16320	rs62581338	GLUC	Central American	C	T	0.994536	0.005464	-0.04	0.41	9.23E-01	915	0.01719
				Cuban	C	T	0.950067	0.049933	0.31	0.12	8.57E-03	1502	0.06125
				Dominican	C	T	0.958032	0.041968	0.47	0.19	1.30E-02	691	0.07989
				Mexican	C	T	0.995448	0.004552	0.17	0.28	5.46E-01	2856	0.00626
				Puerto Rican	C	T	0.982595	0.017405	0.27	0.23	2.47E-01	1264	0.04242
				South American	C	T	0.985507	0.014493	0.73	0.32	2.21E-02	621	0.01741
No	**POS**	**rsID**	**Trait**	**Ancestry**	**A1**	**A2**	**A1_AF**	**MAF**	**β**	**SE**	***P*-value**	**N**	**missRate**
1	8896	rs202120082	WHR	Central American	G	A	0.923529	0.076471	-0.26	0.11	2.59E-02	1020	0.00000
				Cuban	G	A	0.992524	0.007476	-0.46	0.29	1.09E-01	1739	0.00000
				Dominican	NA	NA	NA	NA	NA	NA	NA	NA	NA
				Mexican	G	A	0.996461	0.003539	-0.46	0.29	1.13E-01	3108	0.00000
				Puerto Rican	NA	NA	NA	NA	NA	NA	NA	NA	NA
				South American	G	A	0.989811	0.010189	-0.93	0.37	1.24E-02	687	0.00000
3	14272	rs2853814	BMI	Central American	NA	NA	NA	NA	NA	NA	NA	NA	NA
				Cuban	C	T	0.998395	0.001605	0.33	0.57	5.64E-01	1869	0.02351
				Dominican	C	T	0.998881	0.001119	-2.63	0.96	6.61E-03	894	0.02826
				Mexican	NA	NA	NA	NA	NA	NA	NA	NA	NA
				Puerto Rican	C	T	0.997762	0.002238	-1.34	0.56	1.79E-02	1787	0.00887
				South American	C	T	0.998628	0.001372	-1.74	0.90	5.22E-02	729	0.00410
4	14353		WHR	Central American	T	C	0.924436	0.075564	-0.26	0.12	2.66E-02	1019	0.00098
				Cuban	T	C	0.992524	0.007476	-0.46	0.29	1.09E-01	1739	0.00000
				Dominican	NA	NA	NA	NA	NA	NA	NA	NA	NA
				Mexican	T	C	0.996461	0.003539	-0.46	0.29	1.13E-01	3108	0.00000
				Puerto Rican	NA	NA	NA	NA	NA	NA	NA	NA	NA
				South American	T	C	0.989811	0.010189	-0.93	0.37	1.24E-02	687	0.00000
5	14584		BMI	Central American	NA	NA	NA	NA	NA	NA	NA	NA	NA
				Cuban	T	C	0.998395	0.001605	0.33	0.57	5.64E-01	1869	0.02351
				Dominican	T	C	0.998881	0.001119	-2.63	0.96	6.61E-03	894	0.02826
				Mexican	NA	NA	NA	NA	NA	NA	NA	NA	NA
				Puerto Rican	T	C	0.997762	0.002238	-1.34	0.56	1.79E-02	1787	0.00887
				South American	T	C	0.998628	0.001372	-1.74	0.90	5.22E-02	729	0.00410

Footnote: *These order numbers match with the ones on Tables 1.a-1.b and Tables S2.a-2.b.

Table S4. GWAS Publications, which Summary Results Were Used for Identifying Significant MT-nDNA Candidate Genes

BMI		
Author(s)	PMID	Publication
Speliotes *et al*, 2010	PMID: 20935630	Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-48 (2010).
Yang *et al*, 2012	PMID: 22982992	Yang, J. *et al.* FTO genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267-72 (2012).
Berndt *et al*, 2013	PMID: 23563607	Berndt, S.I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* **45**, 501-12 (2013).
Randall *et al*, 2013	PMID: 23754948	Randall, J.C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* **9**, e1003500 (2013).
Monda *et al*, 2013	PMID: 23583978	Monda, K.L. *et al.* A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nat Genet* **45**, 690-6 (2013).
Locke *et al*, 2015	PMID: 25673413	Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
Shungin *et al*, 2015	PMID: 25673412	Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187-196 (2015).
Winkler *et al*, 2015	PMID: 26426971	Winkler, T.W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* **11**, e1005378 (2015).
NG *et al*, 2017	PMID: 28430825	Ng, M.C.Y. *et al.* Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. *PLoS Genet* **13**, e1006719 (2017).
Justice *et al*, 2017	PMID: 28443625	Justice, A.E. *et al.* Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun* **8**, 14977 (2017).
WHR		
Heid *et al*, 2010	PMID: 20935629	Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* **42**, 949-60 (2010).

Randall *et al*, 2013	PMID: 23754948	Randall, J.C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* **9**, e1003500 (2013).
Berndt *et al*, 2013	PMID: 23563607	Berndt, S.I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* **45**, 501-12 (2013).
Shungin *et al*, 2015	PMID: 25673412	Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187-196 (2015).
Winkler *et al*, 2015	PMID: 26426971	Winkler, T.W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* **11**, e1005378 (2015).
Justice *et al*, 2017	PMID: 28443625	Justice, A.E. *et al.* Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun* **8**, 14977 (2017).
NG *et al*, 2017	PMID: 28430825	Ng, M.C.Y. *et al.* Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. *PLoS Genet* **13**, e1006719 (2017).
Graph *et al*, 2017	PMID: 28448500	Graff, M. *et al.* Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. *PLoS Genet* **13**, e1006528 (2017).
Glucose		
Saxena *et al*, 2010	PMID:20081857	Saxena, R. *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* **42**, 142-8 (2010).
Dupuis *et al*, 2010	PMID:20081858	Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-16 (2010).
Manning *et al*, 2010	PMID:22581228	Manning, A.K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* **44**, 659-69 (2012).
Scott *et al*, 2012	PMID:22885924	Scott, R.A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* **44**, 991-1005 (2012).
INS		
Dupuis *et al*, 2010	PMID:20081858	Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-16 (2010).
Strawbridge *et al*, 2011	PMID: 21873549	Strawbridge, R.J. et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. Diabetes 60, 2624-34.

Manning *et al*, 2010	PMID:22581228	Manning, A.K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* **44**, 659-69 (2012).
Scott *et al*, 2012	PMID:22885924	Scott, R.A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* **44**, 991-1005 (2012).
Prokopenko *et al*, 2014	PMID:24699409	Prokopenko, I. *et al.* A central role for GRB10 in regulation of islet function in man. *PLoS Genet* **10**, e1004235 (2014).
Walford *et al*, 2016	*PMID: 27416945*	Walford, *et al.* Genome-Wide Association Study of the Modified Stumvoll Insulin Sensitivity Index Identifies BCL2 and FAM19A2 as Novel Insulin Sensitivity Loci. *Diabetes 65, 3200-11 (2016)*
HOMAB		
Dupuis *et al*, 2010	PMID:20081858	Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-16 (2010).
	HOMAIR	
Dupuis *et al*, 2010	PMID:20081858	Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-16 (2010).
	HbA1c	
Soranzo *et al*, 2010	PMID:20858683	Soranzo N. *et al. Common variants at 10 genomic loci influence hemoglobin A_1(C) levels via glycemic and nonglycemic pathways. Diabetes 59,3229-39 (2010)*
Wheeler *et al*, 2017	PMID:28898252	Wheeler, E. *et al.* Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med* **14**, e1002383 (2017).

Table S5. Domain Terms (in Pathways and Metabolism-MeSH) Enriched for MT-nDNA Gene Clusters and their Assigned Abstracts in Percent

Estrogen signaling	Antisense	RNA polymorase	Angiogenesis	IGFR-1R	Mitochondrial Carnitine Palmitoyltransferase	PPAR
ESR1, IGF1	BCL2,TNF,VEGFA,IGF1,IGF1R,ESR1,B3GNTL1,FGF2,PPARG,BDNF,RAF1,CDH1,GJA1,NUDT6,POMC,FOXO1,TUT1	TNF,VEGFA,BCL2,ESR1,PPARG,CDH1,IGF1,FGF2,B3GNTL1,IGF1R,BDNF,DNASE1,GJA1,POMC	VEGFA,FGF2	IGF1R,IGF1	CPT1A,TOP1,PPARG	PPARG,NRH1
97,1	24,15,10,5,5,4,4,4,4,3,3,3,2,1,1,1,1	24,12,11,8,7,4,4,4,3,3,3,2,2,1,1	93,6	88,12	66,31,2	98,1
0.0003	0.0005	0.0005	0.001	0.0011	0.0011	0.0015

Tissue distribution	Pentose phosphate pathway	Glycolysis	Cytric Acid cycle	Carbohydrate metabolism	Metabolome	Lipogenesis
POMC,B3GNTL1,ESR1,TSPO,TNF,IGF1,VEGFA,PPARG,FGF2,BDNF,GJA1,APOE,IGF1R,SLC25A26,AKR1A1,DNASE1,CDH1,MAP2,MC4R,CD46,HK1,BCL2,SLC2A2,BCKDK	G6PD,PFKM,HK1	HK1,PFKM,GCK,G6PD	OGDH,HK1,G6PD	HK1,PFKM,G6PD,GCK,B3GNTL1,AKR1A1,FOXO1,PPARG,SLC2A2,OGDH	HK1,CPS1,LPCAT3,CPT1A,G6PD,PPARG,OGDH,CRTC1,PFKM,CYP17A1,MTCH2,TNF,DNLZ,FADS1,IGF1,LRPPRC,PEMT,GCK,ATXN1,APOE	PPARG,NR1H3,CPT1A,FOXO1
14,10,8,7,5,5,4,3,3,3,3,3,2,2,2,2,1,1,1,1,1,1,1,1,1	89,6,4	59,28,7,3	95,3,1	47,16,11,9,6,3,2,1,1,1	13,12,10,9,8,7,6,4,3,2,2,2,2,2,2,2,2,1,1,1	89,5,2,1
0.0041	0.005	0.0054	0.0071	0.0077	0.0097	0.0126

Table S6. Functional Annotation of 109 MT-nDNA Candidate Genes (**Separate Excel Worksheet**)

Table S7. Thirteen and Fifteen Genes out of 109 MT-nDNA Candidates Associated Respectively with T2D and CVD

No	Gene	Mapping	T2D Association Candidates: Function	Trait	*P*-value	PMID
1	*ATXN2*	12q24.12	It is involved in endocytosis, and modulates mTOR signals, modifying ribosomal translation and mitochondrial function. GWAs indicate that loss-of-function mutations in this gene may be associated with susceptibility to type I diabetes, obesity and hypertension	Glycated hemoglobin levels	1.0E-08	28898252
2	*BCL2*	18q21.33	An integral outer mitochondrial membrane protein that blocks the apoptotic death of some cells such as lymphocytes	Type 2 diabetes	4.0E-08	28869590
3	*CPS1*	2q34	The mitochondrial enzyme encoded by this gene catalyzes synthesis of carbamoyl phosphate from ammonia and bicarbonate	Metabolite levels	3.0E-50	23378610
4	*ERAL1*	17q11.2	It is required for proper assembly of the 28S small mitochondrial ribosomal subunit	Glycated hemoglobin levels	3.0E-11	28898252
5	*FADS1*	11q12.2	Regulates unsaturation of fatty acids through the introduction of double bonds between defined carbons of the fatty acyl chain	Fasting glucose related traits	2.0E-15	20081858
6	*G6PD*	Xq28	Glucose-6-phosphate dehydrogenase produces NADPH, a key electron donor in the defense against oxidizing agents and in reductive biosynthetic reactions	Glycated hemoglobin levels	8.0E-135	28898252
7	*HFE*	6p22.2	Iron regulator, associates with beta2-microglobulin	Glycated hemoglobin levels	5.0E-28	28898252
8	*HK1*	10q22.1	Hexokinases phosphorylate glucose to produce glucose-6-phosphate, the first step in most glucose metabolism pathways	Glycated hemoglobin levels	3.0E-65	28898252
9	*IGF1*	12q23.2	It is similar to insulin in function and structure and is involved in mediating growth and development	HOMA-IR	2.0E-09	20081858
10	*PHB2*	12p13.31	PHB2 plays a central role in p21 upregulation following GGCT knockdown and as such may promote deregulated proliferation of cancer	Glycated hemoglobin levels	4.0E-08	28898252
11	*PPARG*	3p25.2	It is a regulator of adipocyte differentiation.	T2D	2.0E-19	28869590
12	*SLC2A2*	3q26.2	Known also as GLUT2: The encoded protein mediates facilitated bidirectional glucose transport. Because of its low affinity for glucose, it has been suggested as a glucose sensor	HbA1c reduction in T2D	7.0E-14	27500523
13	*WFS1*	4p16.1	Down-regulation of WFS1 in neurons leads to changes in mitochondrial dynamics (inhibited mitochondrial fusion, altered mitochondrial trafficking, and augmented mitophagy), delaying neuronal development	T2D	1.0E-15	28869590

No	Gene	Mapping	CVD Association Candidates: Function	Trait	*P*-value	PMID
1	*APOE*	19q13.32	It is essential for the normal catabolism of triglyceride-rich lipoprotein constituents. Mutations in this gene result in familial dysbetalipoproteinemia, or type III hyperlipoproteinemia (HLP III), in which increased plasma cholesterol and triglycerides are the consequence of impaired clearance of chylomicron and VLDL remnants	LDL	2.0E-286	28334899
2	*ATXN2*	12q24.12	It is involved in endocytosis, and modulates mTOR signals, modifying ribosomal translation and mitochondrial function	Coronary artery disease	9.0E-14	28714975
3	*CD46*	1q32.2	The encoded protein has cofactor activity for inactivation of complement components C3b and C4b by serum factor I, which protects the host cell from damage by complement	Resting heart rate	5.0E-31	27798624
4	*CPS1*	2q34	This mitochondrial enzyme catalyzes synthesis of carbamoyl phosphate from ammonia and bicarbonate. This reaction is the first committed step of the urea cycle, which is important in the removal of excess urea from cells	Fibrinogen	2.0E-11	23969696
5	*ESR1*	6q25.1	The estrogen receptor is a ligand-activated transcription factor composed of several domains important for hormone binding, DNA binding, and	Pulse pressure	2.0E-18	28135244
6	*FADS1*	11q12.2	Desaturase enzymes regulate unsaturation of fatty acids through the introduction of double bonds between defined carbons of the fatty acyl chain	LDL	2.0E-39	24097068
7	*FEN1*	11q12.2	Structure-specific nuclease with 5-flap endonuclease and 5-3 exonuclease activities involved in DNA replication and repair	Trans fatty acid levels	5.0E-13	25646338
8	*FHIT*	3p14.2	The protein encoded by this gene is a P1-P3-bis(5'-adenosyl) triphosphate hydrolase involved in purine metabolism	Major coronary event	3.0E-08	28753643
9	*FUNDC2*	Xq28	Ubiquitous expression in heart (RPKM 26.0)	Thrombosis	7.0E-13	26908601
10	*GJA1*	6q22.31	The encoded protein is the major protein of gap junctions in the heart that are thought to have a crucial role in the synchronized contraction of the heart and in embryonic development.	Resting heart rate	3.0E-17	27798624
11	*HFE*	6p22.2	The protein encoded by this gene is a membrane protein that is similar to MHC class I-type proteins and associates with beta2-microglobulin (beta2M)	Diastolic blood pressure	2.0E-15	21909115

No	Gene	Mapping	CVD Association Candidates: Function	Trait	*P*-value	PMID
12	*HMGCR*	5q13.3	HMG-CoA reductase is the rate-limiting enzyme for cholesterol synthesis and is regulated via a negative feedback mechanism mediated by sterols and non-sterol metabolites derived from mevalonate, the product of the reaction catalyzed by reductase	LDL	3.0E-95	28334899
13	*MAPT*	17q21.31	Promotes microtubule assembly and stability, and might be involved in the establishment and maintenance of neuronal polarity	QRS complex (12-leadsum)	2.0E-14	27659466
14	*SFXN2*	10q24.32	Potential iron transporter	Carotid plaque burden	2.0E-08	28282560
15	*STARD13*	13q13.1	It may be involved in regulation of cytoskeletal reorganization, cell proliferation, and cell motility, and acts as a tumor suppressor in hepatoma cells	Intracranial aneurysm	3.0E-09	20364137
16	*SUGCT*	7p14.1	Catalyzes the succinyl-CoA-dependent conversion of glutarate to glutaryl-CoA	Pulse pressure	7.0E-12	28135244
17	*TOMM40*	19q13.32	It is the channel-forming subunit of the translocase of the mitochondrial outer membrane	LDL	2.0E-19	19060911
18	*TOP1*	20q12	Topoisomerases are ubiquitously expressed enzymes that overcome topological problems in genomic DNA, which can result from DNA replication, transcription and repair	LDL	2.0E-34	28334899

Table S8. Several Variants of MT-nDNA Candidates Are eSNV, Affecting Expression of Some of Their Corresponding Genes or Other cis-Located Genes.

No	eSNV	mainGene	TargetGene	AssociatedGWASTrait	eSNV Tissue	eSNVType
1	rs622798	**AKR1A1**	**TESK2**	BMI	Blood	1
19	rs10494221	**WARS2**	**RP11-418J17.1**	WHR	Esophagus muscularis	2
41	rs17369123	**DNM3**	**SUCO**	WHR	Pancreas	2
45	rs6657476	**CD46**	**CD46**	HbA1c	Esophagus muscularis	1
60	rs933994	**CYP27A1**	**RP11-459I19.1**	BMI	Blood	1
78	rs11715915	**AMT**	**RBM6**	GLUC	Skin	2
78	rs11715915	**AMT**	**MST1**	GLUC	Thyroid	1
84	rs10012946	**WFS1**	**WFS1**	GLUC	Skin	1
110	rs9469886	**UHRF1BP1**	**SNRPC**	BMI	Esophagus muscularis	1
140	rs3734264	**UHRF1BP1**	**UHRF1BP1**	BMI	Artery tibial	1
149	rs1799884	**GCK**	**GCK**	GLUC,HOMA-B,HbA1c	Thyroid	2
151	rs3829109	**DNLZ**	**CARD9**	GLUC	Blood	2
157	rs1475644	**SFXN2**	**C10orf32**	WHR	Artery tibial	2
159	rs1004467	**CYP17A1**	**NT5C2**	BMI	Testis	2
164	rs7098825	**AS3MT**	**CYP17A1-AS1**	BMI	Testis	1
164	rs7098825	**AS3MT**	**MARCKSL1P1**	BMI	Adipose subcutaneous	1
165	rs7085104	**AS3MT**	**RP11-724N1.1**	BMI	Brain Cerrebellum	1
166	rs3740390	**AS3MT**	**NT5C2**	BMI	Testis	2
173	rs11030108	**BDNF**	**LIN7C**	BMI	Artery tibial	2
177	rs12573978	**AMBRA1**	**ATG13**	INS	Thyroid	1
180	rs10838681	**NR1H3**	**ATG13**	GLUC	Thyroid	1
180	rs10838681	**NR1H3**	**MADD**	INS	Brain Cerrebellum	2
184	rs7120118	**NR1H3**	**ACP2**	GLUC	Skin	1
184	rs7120118	**NR1H3**	**NR1H3**	GLUC	Testis	1
184	rs7120118	**NR1H3**	**MADD**	GLUC	Brain Cerrebellum	2
186	rs3817335	**MTCH2**	**C1QTNF4**	BMI,INS	Adipose subcutaneous	1
186	rs3817335	**MTCH2**	**SLC39A13**	BMI,INS	Artery tibial	1
187	rs4752856	**MTCH2**	**PSMC3**	BMI,INS	Adipose subcutaneous	1
190	rs7118178	**MTCH2**	**CELF1**	BMI,GLUC,INS	Nerve tibial	1
193	rs174538	**FEN1**	**TMEM258**	GLUC	Fibroblast	1
199	rs174549	**FADS1**	**TMEM258**	GLUC,HbA1c	Fibroblast	1
201	rs174555	**FADS1**	**FADS1**	GLUC,HbA1c	Blood	1
202	rs174556	**FADS1**	**FADS3**	GLUC	Brain Cerrebellum	1
205	rs7124057	**EEF1G**	**EML3**	WHR	Esophagus muscularis	1
206	rs11231150	**TUT1**	**MTA2**	WHR	Skin	1
206	rs11231150	**TUT1**	**EEF1G**	WHR	Artery tibial	1
226	rs740862	**DNASE1**	**TRAP1**	BMI	Muscle skeletal	2
227	rs1053874	**DNASE1**	**CLUAP1**	BMI	Skin	2
227	rs1053874	**DNASE1**	**NLRC3**	BMI	Artery tibial	2

Table S8 (Cont.). Several Variants of MT-nDNA Candidates Are eSNV, Affect Expression of Some of Their Corresponding Genes or Other cis-Located Genes.

No	eSNV	mainGene	TargetGene	AssociatedGWASTrait	eSNV Tissue	eSNVType
236	rs10499	**ATP2A1**	**RP11-1348G14.4**	BMI	Artery tibial	1
237	rs14235	**BCKDK**	**KAT8**	BMI	Skin	1
237	rs14235	**BCKDK**	**VKORC1**	BMI	Liver	1
237	rs14235	**BCKDK**	**ZNF668**	BMI	Blood	1
237	rs14235	**BCKDK**	**PRSS53**	BMI	Liver	1
247	rs242559	**MAPT**	**PROCA1**	HbA1c	Thyroid	2
247	rs2242345	**ERAL1**	**TRAF4**	HbA1c	Fibroblast	1
248	rs6803	**ERAL1**	**TLCD1**	HbA1c	Thyroid	2
261	rs3603	**B3GNTL1**	**RAB40B**	HbA1c	Muscle skeletal	2
271	rs7222773	**B3GNTL1**	**TBCD**	HbA1c	Pancreas	2
273	rs9906163	**B3GNTL1**	**FN3KRP**	HbA1c	Skin	2
289	rs2510344	**NPC1**	**C18orf8**	BMI	Skin	1
304	rs11663558	**NPC1**	**NPC1**	BMI	Adipose subcutaneous	1
326	rs2425056	**UQCC1**	**UQCC1**	WHR	Fibroblast	1
366	rs989711	**TMLHE**	**TMLHE**	HbA1c	Muscle skeletal	1

Notes: No – order number in Supplementary Table 8 matches with order number of eSNVs in Table S10; eSNV – expression variant that affect expression of Target Gene; mainGene is the gene we anchor the eSNV; eSNVs selected are also associated with any of the 7 studied traits (BMI/WHR/GLUC/INS/HOMA-B/HOMA-IR/HbA1c/; the eSNV affect target gene is specific tissue (eSNV Tissue), but there are more tissues where the same eSNV affects the target gene expression; eSNVType is annotated as 1, when the finding is as 'lead' or in high LD with 'lead' regulator (LD r2≥0.80), when annotated as 2 then the finding may represent a 'secondary' regulator independent from the primary regulator (LD r2<0.20). We observed that 'secondary' eSNVs were with larger *P*-values than the 'lead" SNVs, thus will relatively lower contributions to transcription regulation compared to 'lead' eSNVs.

Table S9. Transcription Regulation Evidence on 588 SNVs of 109 MT-nDNA Genes Using RegulomeDB and HaploReg (**Separate Excel Worksheet**)

Table S10. Transcription Regulation of MT-nDNA candidate Variants Using GTEx Software (**Separate Excel worksheet**)

Table S11. Fifteen Genes out of 109 MT-nDNA Candidates Ranked with Highest PageRank Score for PPI

No	Gene	# Interacting Proteins	Mapping	Function	Trait	*P*-value*
1	*WWOX*	259	16q23.1-q23.2	Short-chain dehydrogenases/reductase	WHR	9.99E-07
2	*YWHAE*	432	17p13.3	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein epsilon. It interacts with CDC25 phosphatases, RAF1 and IRS1 proteins, with role in diverse biochemical activities related to signal transduction, such as cell division and regulation of insulin sensitivity	BMI	6.27E-07
3	*RPTOR*	120	17q25.3	A component of a signaling pathway that regulates cell growth in response to nutrient and insulin levels	BMI	2.49E-09
4	*TRIM54*	129	2p23.3	Regulates titin kinase and microtubule-dependent signal pathways in striated muscles	Glucose	3.09E-09
5	*PHB2*	151	12p13.31	Prohibitin-2 plays a central role in p21 upregulation following GGCT knockdown and as such may promote deregulated proliferation of cancer cells by suppressing p21 (*CDKN1A*)[54]	HbA1c	4.44E-08
6	*AMBRA1*	95	11p11.2	Functional deficiency of *Ambra1* in mouse embryos leads to uncontrolled cell proliferation, in addition to severe autophagy impairment[55,56], glucose and insulin	glucose and insulin	4.29E-08, 4.91E-18
7	*RAF1*	194	3p25.2	The RAF1 protein can phosphorylate to activate the dual specificity protein kinases MEK1 and MEK2, which in turn phosphorylate to activate the serine/threonine specific protein kinases, ERK1 and ERK2. Activated ERKs are pleiotropic effectors of cell physiology and play an important role in the control of gene expression involved in the cell division cycle, apoptosis, cell differentiation and cell migration	WHR	3.60E-10
8	*TOP1*	173	20q12	Controls and alters the topologic states of DNA during transcription	WHR	4.11E-07
9	*TOMM40*	126	19q13.32	Is essential for import of protein precursors into mitochondria	BMI	1.25E-08

No	Gene	# Interacting Proteins	Mapping	Function	Trait	*P*-value*
10	*STARD13*	43	13q13.1-q13.2	It may be involved in regulation of cytoskeletal reorganization, cell proliferation, and cell motility	BMI	7.25E-08
11	*ESR1*	771	6q25.1-q25.2	It is an estrogen receptor, a ligand-activated transcription factor	BMI	4.41E-09
12	*ATXN1*	262	6p22.3	It is associated with spinocerebellar ataxia type 1	HbA1c	5.52E-07
13	*LRPPRC*	113	2p21	it may play a role in cytoskeletal organization, vesicular transport, or in transcriptional regulation of both nuclear and mitochondrial genes. The protein localizes primarily to mitochondria	WHR	8.52E-08
14	*MTCH2*	87	11p11.2	It has a regulatory role in adipocyte differentiation	BMI, glucose and insulin	5.12E-08, 3.84E-14, and 2.16E-29
15	*TRAF3*	121	14q32.32	It is important for the activation of the immune response	BMI	2.82E-07
*P-values are sourced from Table 4 in the main text of the manuscript						

Table S12. A Comparison of 109 MT-nDNA Candidate Genes Detected with *P*<1E-06, and Additional SNVs for the Same Genes from GWAS-Catalog Conditional They Pass *P*≤5E-08 (**Separate Excel Worksheet**)

Table S13. A Comparison of mtDNA Association Results of Three Papers with Our Overlapping Results for Different Traits (**Separate Excel Worksheet**)

Table S14. Summary of Chips Used for Studying mtDNA Variants per Cohort and Summary Statistics for Imputation.

No	Study/Cohort	Ancestry	mtDNA Genotyping Array	Number of mtDNA SNPs genotyped	SNPs used for imputation	SNPs imputed	Imputation Accuracy (%)	SNPs after fcGENE
1	AIRWAVE	EA	Illumina HumanCoreExome	401	258	3399	98.9	3653
2	ALSPAC	EA	Illumina HumanHap550	33	33	3643	70.2	3641
3	ARIC	AA	Affy 6.0	111	60	3622	99.3	450
4	ARIC	EA	Affy 6.0	109	60	3622	99.4	205
5	DANISH-60 YEARS	EA	Illumina CoreExome chip	393	285	3655	98.6	549
6	DANISH-ADDITION PRO	EA	Illumina CoreExome chip	392	284	3655	98.7	595
7	DANISH-COGEN	EA	Illumina CoreExome chip	371	284	3656	98.9	672
8	DANISH-GOYA_CASES	EA	Illumina Human660W-quad	58	58	3658	68.3	3041
9	DANISH-GOYA_CONTROLS	EA	Illumina Human660W-quad	58	58	3658	68.3	3041
10	DANISH-HEALTH06	EA	Illumina Human Exome chip	226	194	3633	98.4	522
11	DANISH-INTER99	EA	Illumina Human Exome chip	226	200	3633	98.2	552
12	DANISH-VEJLE	EA	Illumina Human Exome chip	226	199	3633	98.4	548
13	EPIC-INTERACT-COREEXOME	EA	Illumina Human CoreExome-12v1-0_B	398	182	3655	99.8	780
14	EPIC-INTERACT-Ill660W	EA	Illumina 660w quad chip	136	135	3637	99.8	712
15	EPIC-NORFOLK	EA	Affymetrix Axiom UKBiobank	264	257	3674	99.0	928
16	FAMHS	EA	Illumina Human Exome 12v1.0 BeadChip	172	168	3631	99.9	390
17	FENLAND	EA	Affymetrix Axiom UKBiobank	268	261	3674	99.1	932
18	FHS	EA	Exome Chip, Customized Chip	199	199	3636	99.9	639
19	GENESTAR	AA	Illumina Human1M_v1C BeadChip	162	162	3642	99.8	714
20	GENESTAR	EA	Illumina Human1M_v1C BeadChip	162	162	3642	99.8	492
21	GENOA	AA	Affy 6.0, Illumina 1M-Duo	90	63	3621	99.5	3617
22	GENOA	EA	Affy 6.0, Illumina 660W, Illumina 1M-Duo	56	45	3620	99.3	3615
23	GOLDN	EA	Affy 6.0	110	110	3710	77.8	55
24	HCHS-SOL	HA	Illumina Omni 2.5M chip + 109,571 custom SNPs	392	314	3667	99.3	757
25	HYPERGEN	AA	Affy 6.0	110	44	3619	86.2	3616
26	HYPERGEN	EA	Illumina Cardio-MetaboChip	135	52	3622	98.4	3618
27	LBC-1921	EA	Exome chip	20	20	3621	61.2	2102
28	LBC-1936	EA	Exome chip	20	20	3621	60.6	2194
29	LLFS	EA	Illumina Omni chip	253	143	3653	99.7	453
30	MESA	AA	Affy 6.0	226	172	3630	98.5	505
31	MESA	EA	Affy 6.0	226	172	3630	99.6	512
32	MESA	ASA	Affy 6.0	226	172	3630	99.3	535
33	MESA	HA	Affy 6.0	226	172	3630	99.0	534
34	NEO	EA	Illumina HumanCoreExome chip v1	306	262	3660	97.7	637
35	OOA	EA	Illumina CoreExome chip	364	350	3623	98.9	3623
36	PELOTAS	EA	Illumina HumanOmni2.5-8v1	220	125	3678	99.7	356
37	PELOTAS	Brazilian	Illumina HumanOmni2.5-8v1	220	125	3678	99.7	356
38	ROTTERDAM STUDY II	EA	illumina 550K (duo)	73	73	3638	97.0	3630
39	ROTTERDAM STUDY III	EA	Illumina 610k (quad)	110	110	3636	97.0	3630
40	SHIP	EA	Illumina Infinium Human Exome BeadChip v1.0	163	163	3630	99.8	3625
41	TUDR	ASA	Illumina Exome Chip	198	198	3632	99.1	547
42	WGHS	EA	Illumina HumanExome Beadchip v.1.1	204	203	3643	99.2	836
43	WHI	AA	Affy 6.0	119	119	3678	91.7	525
44	WHI	HA	Affy 6.0	119	119	3678	91.0	426
45	WUHHD	EA	Affy 6.0	167	95	3642	95.5	3640

Table S15. Illumina Infinium Omni Express Exome v1.1

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs41531144	Non-coding	T217C					
rs41323649	Non-coding	G228A					
2263307	Non-coding	C285T				Elderly fibroblasts	
rs41528348	Non-coding	C295T				POLG/MNGIE muscle; Glioblastoma	
2216184	Non-coding	C418T					
rs28625645	Non-coding	T489C				Ovarian carcinoma;	
rs3901846	Non-coding	G499A				Thyroid & prostate tumors	
2216185	*tRNA-Phe*	A606G			Myoglobinuria		
2216186	*tRNA-Phe*	T629C					
2216191	*12S rRNA*	T710C				Colorectal tumor	
2216198	*12S rRNA*	T921C			Possibly LVNC-associated		
2216200	*12S rRNA*	G951A					
2216201	*12S rRNA*	T961C			DEAF; Possibly LVNC-associated		
2216202	*12S rRNA*	T1005C			DEAF		
2216203	*12S rRNA*	G1018A					
2216204	*12S rRNA*	C1048T					
2216208	*12S rRNA*	T1243C				Pancreatic cancer cell line	
2216209	*12S rRNA*	A1382C					
2216210	*12S rRNA*	G1393A					
2216211	*12S rRNA*	T1406C				Pancreatic cancer cell line	
2216212	*12S rRNA*	T1413C					
2216213	*12S rRNA*	A1438G					
2216214	*12S rRNA*	G1442A					
2216216	*16S rRNA*	G1664A					
2216217	*16S rRNA*	T1694C					
2216218	*16S rRNA*	T1700C					
2216219	*16S rRNA*	C1703T					
2216220	*16S rRNA*	C1706T					
2216221	*16S rRNA*	G1709A					

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2216222	*16S rRNA*	T1717C					
rs3928305	*16S rRNA*	G1719A					
2216228	*16S rRNA*	A1811G				Head/neck tumor	
2216231	*16S rRNA*	A1842G					
2216232	*16S rRNA*	G1888A					
2216234	*16S rRNA*	G2056A				Bladder tumor; Acute leukemia platelets, leukocytes, bone marrow	
2216238	*16S rRNA*	C2218T					
2216239	*16S rRNA*	C2259T					
2216240	*16S rRNA*	C2283T					
2216241	*16S rRNA*	C2332T					
2216242	*16S rRNA*	T2352C			Possibly LVNC-associated		
2216243	*16S rRNA*	A2358G					
2216244	*16S rRNA*	T2416C					
2216245	*16S rRNA*	C2417G					
2216249	*16S rRNA*	A2706G					
2216250	*16S rRNA*	G2758A					
2216251	*16S rRNA*	A2768G					
2216252	*16S rRNA*	C2772T					
2216254	*16S rRNA*	G2831A					
2216255	*16S rRNA*	T2885C					
2216256	*16S rRNA*	G3010A			Cyclic vomiting syndrome with migraine		
2216257	*16S rRNA*	T3027C					
2216258	*16S rRNA*	C3116T					
2216259	*16S rRNA*	T3197C					
2216260	*16S rRNA*	T3200A					
2216261	*16S rRNA*	C3206T					
2216262	*16S rRNA*	C3210T					
rs2853516	*ND1*	G3316A	Nonsynonymous	A4T	Diabetes; LHON; PEO		Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2216265	*ND1*	T3338C	Nonsynonymous	V11A			Benign
2216266	*ND1*	T3394C	Nonsynonymous	Y30H	LHON; Diabetes; CPT deficiency; High altitude adaptation	Acute leukemia platelets, leukocytes, & bone marrow	Benign
2216267	*ND1*	T3398C	Nonsynonymous	M31T	DMDF+HCM; GDM; Possibly LVNC cardiomyopathy-associated		Benign
2216268	*ND1*	A3434G	Nonsynonymous	Y43C		Prostate tumor	Benign
2216269	*ND1*	C3497T	Nonsynonymous	A64V	LHON		Benign
2216271	*ND1*	A3547G	Nonsynonymous	I81V			Benign
2216272	*ND1*	C3571T	Nonsynonymous	L89F			Probably damaging
2263308	*ND1*	G3736A	Nonsynonymous	V144I	LHON		Benign
2216273	*ND1*	C3746T	Nonsynonymous	A147V			Benign
2216274	*ND1*	T3866C	Nonsynonymous	I187T	LHON + limb claudication		Benign
2216277	*ND1*	A4021G	Nonsynonymous	T239A			Benign
2216280	*ND1*	A4093G	Nonsynonymous	T263A			Benign
2216281	*ND1*	A4123G	Nonsynonymous	I273V			Benign
2216282	*ND1*	A4129G	Nonsynonymous	T275A			Benign
2216284	*tRNA-Ile*	C4312T				Thyroid tumor	
rs41456348	*tRNA-Gly*	T4336C			AD; PD; Hearing loss & migraine		
2216286	*tRNA-Gln*	T4363C			Possibly associated with DEAF + RP + developmental delay; Hypertension		
2216294	*ND2*	A4732G	Nonsynonymous	N88S			Benign
2263338	*ND2*	A4917G	Nonsynonymous	N150D	LHON; Insulin resistance; AMD; NRTI-PN; Haplogroup T marker		Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2216303	*ND2*	C5331A	Nonsynonymous	L288I			Benign
rs3021088	*ND2*	G5460A	Nonsynonymous	A331T	AD; PD		Benign
2216317	*tRNA-Cys*	T5814C			Mitochondrial encephalopathy; Haplogroup H2b marker	MNGIE tissues	
2216322	*CO1*	T6253C	Nonsynonymous	M117T	Prostate cancer; Enriched in POAG cohort		Benign
2216323	*CO1*	G6261A	Nonsynonymous	A120T	Prostate cancer; LHON		Probably damaging
2216324	*CO1*	G6267A	Nonsynonymous	A122T	Prostate cancer	Pancreatic cancer cell line	Benign
2216325	*CO1*	G6480A	Nonsynonymous	V193I	Prostate cancer; Enriched in POAG cohort		Benign
2263312	*CO1*	A6663G	Nonsynonymous	I254V	Prostate cancer		Benign
2263313	*CO1*	A6982G	Nonsynonymous	N360S			Probably damaging
2216332	*tRNA-Ser*	C7476T				Thyroid hyperplasia	
2263318	*ATP8*	A8411C	Nonsynonymous	M16L			Benign
2216340	*ATP8*	C8414T	Nonsynonymous	L17F	Longevity		Probably damaging
2216341	*ATP8*	C8417T	Nonsynonymous	L18F			Probably Damaging
2216342	*ATP8*	A8460G	Nonsynonymous	N32S			Possibly Damaging
2216343	*ATP8*	C8478T	Nonsynonymous	S38L			Benign
2263319	*ATP6*	C8684T	Nonsynonymous	T53I			Benign
2263340	*ATP6*	A8701G	Nonsynonymous	T59A		Thyroid tumors	Benign
2263320	*ATP6*	G8857A	Nonsynonymous	G111S			Benign
2216361	*ATP6*	G8896A	Nonsynonymous	A124T			Benign
2216366	*ATP6*	G9053A	Nonsynonymous	S176N			Benign
2216368	*ATP6*	T9098C	Nonsynonymous	I191T	Predisposition to anti-retroviral mitochondrial disease		Probably damaging
rs2853825	*CO3*	G9477A	Nonsynonymous	V91I		Thyroid tumor	Benign
2263321	*CO3*	G9804A	Nonsynonymous	A200T	LHON	Pancreatic cancer cell line	Benign
2216376	*CO3*	A9855G	Nonsynonymous	I217V			Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2216377	*CO3*	T9903C	Nonsynonymous	F233L			Probably damaging
2216378	*CO3*	G9966A	Nonsynonymous	V254I			Benign
2216379	*tRNA-Gly*	T10007C					
2216380	*tRNA-Gly*	T10031C					
2263322	*tRNA-Gly*	T10034C					
2263323	*ND3*	T10084C	Nonsynonymous	I9T			Benign
2216384	*ND3*	A10086G	Nonsynonymous	N10D	Hypertensive end-stage renal disease		Probably damaging
2216385	*ND3*	G10143A	Nonsynonymous	G29S			Benign
2216386	*ND3*	T10321C	Nonsynonymous	V88A		Bladder tumor	Benign
2216387	*ND3*	T10345C	Nonsynonymous	I96T			Benign
rs2853826	*ND3*	A10398G	Nonsynonymous	T114A	PD protective factor; Longevity; Altered cell pH; Metabolic syndrome; Breast cancer risk; Haplogroup IJK marker		Benign
2216389	*tRNA-Arg*	T10410C					
rs28358279	*tRNA-Arg*	T10463C				Endometrium tumor	
2216391	*ND4L*	T10609C	Nonsynonymous	M47T			Benign
2216394	*ND4*	T11025C	Nonsynonymous	L89P			Benign
2216396	*ND4*	A11172G	Nonsynonymous	N138S			Probably damaging
2216397	*ND4*	C11177T	Nonsynonymous	P140S			Probably damaging
2216398	*ND4*	T11204C	Nonsynonymous	F149L			Benign
2216399	*ND4*	T11253C	Nonsynonymous	I165T	LHON; PD		Benign
2216400	*ND4*	G11696A	Nonsynonymous	V313I	LHON; LDYT; DEAF; Hypertension helper mutation		Benign
2216401	*ND4*	G11963A	Nonsynonymous	V402I			Benign
2216402	*ND4*	G11969A	Nonsynonymous	A404T			Benign
2216403	*ND4*	T11984C	Nonsynonymous	Y409H			Probably damaging
2216406	*tRNA-Ser*	G12236A			DEAF	Thyroid tumor	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2216407	*tRNA-Ser*	A12248G					
rs2853498	*tRNA-Leu2*	A12308G			CPEO; Stroke; CM; Breast, renal, and prostate cancer risk; Altered brain pH; sCJD; Haplogroup U marker	Endometrium control tissue; Lung and prostate tumors	
2216414	*ND5*	A12397G	Nonsynonymous	T21A	Early onset PD		Unknown
2216417	*ND5*	G12454A	Nonsynonymous	V40I			Benign
2216418	*ND5*	C12542T	Nonsynonymous	A69V			Possibly damaging
2263325	*ND5*	G12820A	Nonsynonymous	A162T			Benign
2263326	*ND5*	A12937G	Nonsynonymous	M201V			Benign
2216420	*ND5*	G12940A	Nonsynonymous	A202T			Benign
2216421	*ND5*	A12950G	Nonsynonymous	N205S			Benign
2216422	*ND5*	A13105G	Nonsynonymous	I257V			Benign
2216423	*ND5*	G13135A	Nonsynonymous	A267T	Possible HCM susceptibility		Benign
2216424	*ND5*	A13276G	Nonsynonymous	M314V			Benign
2216425	*ND5*	G13477A	Nonsynonymous	A381T			Benign
rs28359178	*ND5*	G13708A	Nonsynonymous	A458T	LHON; Increased MS risk; Higher frequency in PD-ADS; Haplogroup J marker	Acute leukemia platelets, leukocytes, & bone marrow; Breast tumor	Benign
2263327	*ND5*	A13780G	Nonsynonymous	I482V			Benign
2216429	*ND5*	T13789C	Nonsynonymous	Y485H			Probably damaging
2216431	*ND5*	C13880A	Nonsynonymous	S515Y			Possibly damaging
2216432	*ND5*	T13886C	Nonsynonymous	L517P			Benign
2216433	*ND5*	C13924T	Nonsynonymous	P530S			Probably damaging
2216434	*ND5*	A13933G	Nonsynonymous	T533A			Benign
2263328	*ND5*	C13934T	Nonsynonymous	T533M			Benign
2216436	*ND5*	A13942G	Nonsynonymous	T536A			Benign
2216437	Non-Coding	G13958C	Nonsynonymous	G541A			Possibly damaging

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2263329	*ND5*	A13966G	Nonsynonymous	T544A			Benign
2216439	*ND5*	C13981T	Nonsynonymous	P549S			Benign
2216440	*ND5*	T14000A	Nonsynonymous	L555Q			Probably damaging
2216441	*ND5*	A14053G	Nonsynonymous	T573A		Prostate tumor	Benign
2216442	*ND5*	A14059G	Nonsynonymous	I575V			Benign
2216443	*ND5*	A14148G	Synonymous	Term604Term			
2216444	*ND6*	T14178C	Nonsynonymous	I166V			Benign
2216445	*ND6*	T14180C	Nonsynonymous	Y165C			Probably damaging
2216446	*ND6*	G14258A	Nonsynonymous	P139L			Benign
2216447	*ND6*	T14318C	Nonsynonymous	N119S			Benign
2216448	*ND6*	T14502C	Nonsynonymous	I58V	LHON		Benign
2263330	*tRNA-Glu*	A14687G			Mitochondrial myopathy with respiratory failure	Mitochondrial myopathy with respiratory failure	
rs3135031	*CYB*	C14766T	Nonsynonymous	T7I			Benign
rs28357681	*CYB*	T14798C	Nonsynonymous	F18L		Glioblastoma	Benign
2216456	*CYB*	G14861A	Nonsynonymous	A39T			Benign
2216457	*CYB*	A14927G	Nonsynonymous	T61A			Benign
2216458	*CYB*	A14978G	Nonsynonymous	I78V			Benign
2216459	*CYB*	T14979C	Nonsynonymous	I78T			Benign
2216460	*CYB*	A15038G	Nonsynonymous	I98V			Possibly damaging
2216461	*CYB*	T15071C	Nonsynonymous	Y109H			Benign
2216462	*CYB*	T15074C	Nonsynonymous	S110P			Benign
2216463	*CYB*	G15077A	Nonsynonymous	E111K	DEAF		Probably damaging
2216464	*CYB*	G15110A	Nonsynonymous	A122T			Benign
2216465	*CYB*	T15204C	Nonsynonymous	I153T			Benign
2263332	*CYB*	G15119A	Nonsynonymous	A125T			Benign
rs2853506	*CYB*	A15218G	Nonsynonymous	T158A			Possibly damaging
2216467	*CYB*	A15236G	Nonsynonymous	I164V			Benign
2216468	*CYB*	G15257A	Nonsynonymous	D171N	LHON; Haplogroup J2 marker		Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2263333	*CYB*	C15263T	Nonsynonymous	P173S			Benign
2216469	*CYB*	A15311G	Nonsynonymous	I189V			Benign
2216470	*CYB*	G15314A	Nonsynonymous	A190T			Benign
2216471	*CYB*	G15323A	Nonsynonymous	A193T			Benign
2216472	*CYB*	A15326G	Nonsynonymous	T194A			Benign
2216473	*CYB*	C15381T	Nonsynonymous	T212I			Benign
2216474	*CYB*	C15402T	Nonsynonymous	T219I			Possibly damaging
2216475	*CYB*	G15431A	Nonsynonymous	A229T			Benign
rs3088309	*CYB*	C15452A	Nonsynonymous	L236I			Benign
2216478	*CYB*	T15479C	Nonsynonymous	F245L			Benign
2216479	*CYB*	G15497A	Nonsynonymous	G251S	EXIT; Obesity		Benign
2216481	*CYB*	A15662G	Nonsynonymous	I306V	Complex mitochondriopathy-associated	Breast tumor	Benign
2216482	*CYB*	T15672C	Nonsynonymous	M309T		Breast tumor	Benign
2263334	*CYB*	T15693C	Nonsynonymous	M316T	Possibly LVNC cardiomyopathy-associated	Breast tumor	Benign
rs41337244	*CYB*	A15758G	Nonsynonymous	I338V			Benign
2216486	*CYB*	A15824G	Nonsynonymous	T360A		Breast tumor	Benign
2216487	*CYB*	C15849T	Nonsynonymous	T368I		Breast tumor	Benign
2216489	*CYB*	A15860G	Nonsynonymous	I372V			Benign
2263335	*tRNA-Thr*	T15889C					
2263341	*tRNA-Thr*	C15904T					
2216491	*tRNA-Thr*	T15905C					
2216492	*tRNA-Thr*	A15907G					
rs2853510	*tRNA-Thr*	A15924G			LIMM		
2263337	*tRNA-Thr*	G15928A			Multiple sclerosis; Idiopathic repeat miscarriage; AD protection		
2216495	*tRNA-Thr*	C15939T					

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2263336	*tRNA-Thr*	T15940C					
2216496	*tRNA-Thr*	T15941C					
2216497	*tRNA-Thr*	T15942C			Possibly LVNC-associated		
2216498	*tRNA-Thr*	C15946T					
2216499	*tRNA-Thr*	A15951G			LHON modulator		
2216500	*tRNA-Pro*	C15978T					
2216501	*tRNA-Pro*	T16017C					
rs41378955	Non-coding	G16390A			POAG-potential for association	Breast and ovarian tumors	

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; POLG = Mitochondrial DNA polymerase ɣ; MNGIE = Mitochondrial neurogastrointestinal encephalopathy; LVNC = Left ventricular non-compaction cardiomyopathy; DEAF = Deafness; LHON = Leber Hereditary optic neuropathy; PEO = Progressive external ophthalmoplegia; CPT = Carnitine palmityltransferase deficiency; DMDF = Diabetes mellitus and deafness; HCM = Hypertrophic cardiomyopathy; GDM = Gestational diabetes mellitus; AD = Alzheimer's disease; PD = Parkinson's disease; RP = Retinitis pigmentosa; AMD = Age-related macular degeneration; POAG = Primary open angle glaucoma; LDYT = Leber hereditary optic neuropathy and dystonia; CM = Cardiomyopathy; sCJD = Sporadic Creutzfeldt-Jaokob disease; MS = Multiple sclerosis; PD-ADS = Acquired demyelinating syndromes; HCM = Hypertrophic cardiomyopathy; EXIT = Exercise intolerance; LIMM = Lethal infantile mitochondrial myopathy

Table S16. Infinium Human Exome v1.0

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs28625645	Non-coding	T89C					
rs62581312	Non-coding	C150T			Longevity; Cervical carcinoma; HPV infection risk	Elderly fibroblasts; Leukocytes, lung, thyroid, prostate tumors	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs41531144	Non-coding	T217C					
rs41323649	Non-coding	G228A					
rs41528348	Non-coding	C295T				POLG &MNGIE muscle; Glioblastoma	
rs3901846	Non-coding	G499A				Thyroid & prostate tumors	
rs3928305	*16S rRNA*	G1719A					
rs2853516	*ND1*	G3316A	Nonsynonymous	A4T	Diabetes; LHON; PEO		Benign
rs41456348	*tRNA-Gln*	T4336C			AD; PD; Hearing loss & migraine		
rs3021088	*ND2*	G5460A	Nonsynonymous	A331T	AD; PD		Benign
rs2853825	*CO3*	G9477A	Nonsynonymous	V91I		Thyroid tumor	Benign
rs2853826	*ND3*	A10398G	Nonsynonymous	T114A	PD protective factor; Longevity; Altered cell pH; Metabolic syndrome; Breast cancer risk; Haplogroup IJK marker	Thyroid tumor	Benign
rs28358279	*tRNA-Arg*	T10463C				Endometrium tumor	
rs2853498	*tRNA-Leu2*	A12308G					
rs28359178	*ND5*	G13708A	Nonsynonymous	A458T	LHON; Increased MS risk; Higher frequency in PD/ADS; Haplogroup J marker	Acute leukemia platelets, leukocytes, & bone marrow; Breast tumor	Benign
rs3135031	*CYB*	C14766T	Nonsynonymous	T7I			Benign
rs28357681	*CYB*	T14798C	Nonsynonymous	F18L		Glioblastoma	Benign
rs2853506	*CYB*	A15218G	Nonsynonymous	T158A			Possibly damaging
rs3088309	*CYB*	C15452A	Nonsynonymous	L236I			Benign
rs41337244	*CYB*	A15758G	Nonsynonymous	I338V			Benign
rs2853510	*tRNA-Thr*	A15924G			LIMM		
rs41378955	Non-coding	G16390A			POAG- potential for association	Breast, ovarian tumor	

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; HPV = Human papillomavirus; POLG = Mitochondrial DNA polymerase ɣ; MNGIE = Mitochondrial neurogastrointestinal encephalopathy; LHON = Leber Hereditary optic neuropathy; PEO = Progressive external ophthalmoplegia; AD = Alzheimer's disease; PD = Parkinson's disease; MS = Multiple sclerosis; LIMM = Lethal infantile mitochondrial myopathy; POAG = Primary open angle glaucoma

Table S17. Infinium Core Exome v1.1

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2010-08-MT-841	Non-coding	T72C				Aging brains; POLG/PEO & control muscle; Normal tissues	
2010-08-MT-981	Non-coding	A93G					
200610-102	Non-coding	T125C				POLG/PEO & control muscle	
2010-08-MT-550	Non-coding	A215G				Esophageal cancer	
rs41531144	Non-coding	T217C					
rs41323649	Non-coding	G228A					
200610-105	Non-coding	T236C					
2263307	Non-coding	C285T				Elderly fibroblasts	
rs41528348	Non-coding	C295T				POLG/MNGIE muscle; Glioblastoma	
2216184	Non-coding	C418T					
rs28625645	Non-coding	T489C				Ovarian carcinoma; Prostate cancer; Thyroid & prostate tumors	
rs3901846	Non-coding	G499A					
2216185	*tRNA-Phe*	A606G			Myoglobinuria		
2216186	*tRNA-Phe*	T629C					

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2216191	*12S rRNA*	T710C				Colorectal tumor	
2010-08-MT-830	*12S rRNA*	T711C					
200610-42	*12S rRNA*	C722T					
2216198	*12S rRNA*	T921C			Possibly LVNC-associated		
2216200	*12S rRNA*	G951A					
2216201	*12S rRNA*	T961C			DEAF; Possibly LVNC-associated		
2216202	*12S rRNA*	T1005C			DEAF		
2216203	*12S rRNA*	G1018A					
2010-08-MT-27	*12S rRNA*	A1041G					
2216204	*12S rRNA*	C1048T					
200610-108	*12S rRNA*	T1107C					
200610-109	*12S rRNA*	T1119C					
2216208	*12S rRNA*	T1243C				Pancreatic cancer cell line	
2216209	*12S rRNA*	A1382C					
2216210	*12S rRNA*	G1393A					
2216211	*12S rRNA*	T1406C				Pancreatic cancer cell line	
2216212	*12S rRNA*	T1413C					
2216213	*12S rRNA*	A1438G					
2216214	*12S rRNA*	G1442A					
2216216	*16S rRNA*	G1664A					
2216217	*16S rRNA*	T1694C					
2216218	*16S rRNA*	T1700C					
2216219	*16S rRNA*	C1703T					
2216220	*16S rRNA*	C1706T					
2216221	*16S rRNA*	G1709A					
2216222	*16S rRNA*	T1717C					
rs3928305	*16S rRNA*	G1719A					

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2010-08-MT-526	*16S rRNA*	C1721T				Acute leukemia platelets, leukocytes, & bone marrow	
2216228	*16S rRNA*	A1811G				Head/neck tumor	
2216231	*16S rRNA*	A1842G					
2216232	*16S rRNA*	G1888A					
2216234	*16S rRNA*	G2056A				Bladder tumor; Acute leukemia; Platelets/leukocytes/bone marrow	
2216238	*16S rRNA*	C2218T					
2216239	*16S rRNA*	C2259T					
2216240	*16S rRNA*	C2283T					
2216241	*16S rRNA*	C2332T					
2216242	*16S rRNA*	T2352C			Possibly LVNC-associated		
2216243	*16S rRNA*	A2358G					
2216244	*16S rRNA*	T2416C					
2216249	*16S rRNA*	A2706G					
2216250	*16S rRNA*	G2758A					
2216251	*16S rRNA*	A2768G					
2216252	*16S rRNA*	C2772T					
2216254	*16S rRNA*	G2831A					
200610-2	*16S rRNA*	A2880G					
2216255	*16S rRNA*	T2885C					
2216256	*16S rRNA*	G3010A			Cyclic vomiting syndrome with migraine		
2216257	*16S rRNA*	T3027C					
2216258	*16S rRNA*	C3116T					
2216259	*16S rRNA*	T3197C					
2216260	*16S rRNA*	T3200A					
2216261	*16S rRNA*	C3206T					
2216262	*16S rRNA*	C3210T					

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs2853516	*ND1*	G3316A	Nonsynonymous	A4T	Diabetes; LHON; PEO		Benign
2216265	*ND1*	T3338C	Nonsynonymous	V11A			Benign
200610-4	*ND1*	A3384G	Synonymous	K26K			
2216266	*ND1*	T3394C	Nonsynonymous	Y30H	LHON; Diabetes; CPT deficiency; High altitude adaptation	Acute leukemia platelets, leukocytes, & bone marrow	Benign
200610-111	*ND1*	T3396C	Synonymous	Y30Y	NSHL; MIDD		
2216267	*ND1*	T3398C	Nonsynonymous	M31T		DMDF+HCM; GDM; Possibly LVNC cardiomyopathy-associated	Benign
2216268	*ND1*	A3434G	Nonsynonymous	Y43C		Prostate tumor	Benign
2216269	*ND1*	C3497T	Nonsynonymous	A64V	LHON		Benign
2216271	*ND1*	A3547G	Nonsynonymous	I81V			Benign
2216272	*ND1*	C3571T	Nonsynonymous	L89F			Probably damaging
200610-112	*ND1*	T3644C	Nonsynonymous	V113A	BD-associated		Benign
200610-113	*ND1*	T3645C	Synonymous	V113V			
2263308	*ND1*	G3736A	Nonsynonymous	V144I	LHON		Benign
2216273	*ND1*	C3746T	Nonsynonymous	A147V			Benign
200610-114	*ND1*	T3826C	Synonymous	L174L			
2216274	*ND1*	T3866C	Nonsynonymous	I187T	LHON + limb claudication		Benign
200610-45	*ND1*	C3921T	Synonymous	S205S			
2216277	*ND1*	A4021G	Nonsynonymous	T239A			Benign
200610-115	*ND1*	T4023C	Synonymous	T239T			
200610-47	*ND1*	C4025T	Nonsynonymous	T240M			Benign
2010-08-MT-655	*ND1*	G4048A	Nonsynonymous	D248N			Benign
2216280	*ND1*	A4093G	Nonsynonymous	T263A			Benign
200610-116	*ND1*	T4117C	Synonymous	L271L			
2216281	*ND1*	A4123G	Nonsynonymous	I273V			Benign
2216282	*ND1*	A4129G	Nonsynonymous	T275A			Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2010-08-MT-664	*ND1*	A4164G	Synonymous	M286M			
200610-117	*ND1*	T4218C	Synonymous	Y304Y			
2216284	*tRNA-Ile*	C4312T				Thyroid tumor	
200610-48	*tRNA-Gln*	C4335T					
rs41456348	*tRNA-Gln*	T4336C			AD; PD; Hearing loss & migraine		
2216286	*tRNA-Gln*	T4363C			Possibly associated with DEAF + RP +developmental delay; Hypertension		
2216291	*ND2*	G4491A	Nonsynonymous	V8I			Benign
200610-49	*ND2*	C4508T	Synonymous	I13I			
2216294	*ND2*	A4732G	Nonsynonymous	N88S			Benign
200610-6	*ND2*	A4833G	Nonsynonymous	T122A	Diabetes helper mutation; AD; PD; Haplogroup G marker		Possibly damaging
2263338	*ND2*	A4917G	Nonsynonymous	N150D	LHON; Insulin resistance; AMD; NRTI-PN; Haplogroup T marker		Benign
200610-120	*ND2*	T5108C	Synonymous	T213T			
200610-7	*ND2*	A5301G	Nonsynonymous	I278V			Benign
2216303	*ND2*	C5331A	Nonsynonymous	L288I			Benign
rs3021088	*ND2*	G5460A	Nonsynonymous	A331T	AD; PD		Benign
2216317	*tRNA-Cys*	T5814C			Mitochondrial encephalopathy; L2b marker	MNGIE tissues	
200610-9	*tRNA-Tyr*	A5833G					
2010-08-MT-773	*CO1*	T5999C	Synonymous	A32A		Pancreatic cancer cell line; Glioblastoma	
2010-08-MT-776	*CO1*	A6047G	Synonymous	L48L		Pancreatic cancer cell line; Glioblastoma	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
200610-51	*CO1*	C6077T	Synonymous	V58V			
200610-121	*CO1*	T6248C	Synonymous	S115S			
2216322	*CO1*	T6253C	Nonsynonymous	M117T	Prostate cancer; Enriched in POAG cohort		Benign
2216323	*CO1*	G6261A	Nonsynonymous	A120T	Prostate cancer; LHON		Probably damaging
2216324	*CO1*	G6267A	Nonsynonymous	A122T	Prostate cancer	Pancreatic cancer cell line	Benign
200610-80	*CO1*	G6285A	Nonsynonymous	V128I	Prostate cancer		Benign
200610-123	*CO1*	T6392C	Synonymous	N163N			
200610-81	*CO1*	G6446A	Synonymous	T181T			
2216325	*CO1*	G6480A	Nonsynonymous	V193I	Prostate cancer; Enriched in POAG cohort		Benign
2263312	*CO1*	A6663G	Nonsynonymous	I254V	Prostate cancer		Benign
200610-82	*CO1*	G6734A	Synonymous	M277M			
2263313	*CO1*	A6982G	Nonsynonymous	N360S			Probably damaging
2010-08-MT-831	*CO1*	T7142C	Synonymous	H413H			
2216332	*tRNA-Ser*	C7476T				Thyroid hyperplasia	
200610-84	*CO2*	G7598A	Nonsynonymous	A5T	Possible LHON helper variant		Benign
2010-08-MT-860	*CO2*	T7684C	Synonymous	L33L			
2010-08-MT-861	*CO2*	G7697A	Nonsynonymous	V38I	Possible HCM susceptibility		Benign
2010-08-MT-874	*CO2*	G7853A	Nonsynonymous	V90I			Benign
200610-86	*CO2*	G7859A	Nonsynonymous	D92N	Progressive encephalomyopathy		Benign
200610-128	*CO2*	T7870C	Synonymous	L95L			
200610-13	*CO2*	A7972G	Synonymous	E129E			
200610-87	*CO2*	G8020A	Synonymous	P145P			
200610-129	*ATP8*	T8404C	Synonymous	I13I			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2216340	*ATP8*	C8414T	Nonsynonymous	L17F	Longevity		Probably damaging
2216341	*ATP8*	C8417T	Nonsynonymous	L18F			Probably damaging
2216342	*ATP8*	A8460G	Nonsynonymous	N32S			Possibly damaging
2216343	*ATP8*	C8478T	Nonsynonymous	S38L			Benign
200610-14	*ATP8*	A8502G	Nonsynonymous	N46S			Possibly damaging
2263319	*ATP6*	C8684T	Nonsynonymous	T53I			Benign
2263340	*ATP6*	A8701G	Nonsynonymous	T59A		Thyroid tumors	Benign
200610-131	*ATP6*	T8793C	Synonymous	P89P			
2263320	*ATP6*	G8857A	Nonsynonymous	G111S			Benign
200610-55	*ATP6*	C8859T	Synonymous	G111G			
2216358	*ATP6*	A8860G	Nonsynonymous	T112A			Benign
2216360	*ATP6*	T8875C	Nonsynonymous	F117L			Benign
2216361	*ATP6*	G8896A	Nonsynonymous	A124T			Benign
2216362	*ATP6*	A8923G	Nonsynonymous	T133A			Probably damaging
200610-18	*ATP6*	A8946G	Synonymous	M140M			
200610-56	*ATP6*	C8964T	Synonymous	T146T			
200610-89	*ATP6*	G9064A	Nonsynonymous	A180T			Benign
2216368	*ATP6*	T9098C	Nonsynonymous	I191T	Predisposition to anti-retroviral mitochondrial disease		Probably damaging
2216369	*ATP6*	T9128C	Nonsynonymous	I201T			Benign
200610-57	*ATP6*	C9140T	Nonsynonymous	A205V			Probably damaging
2216370	*ATP6*	G9142A	Nonsynonymous	V206I			Benign
2010-08-MT-969	*ATP6*	A9150G	Synonymous	L208L			
200610-58	*CO3*	C9458T	Synonymous	I84I			
rs2853825	*CO3*	G9477A	Nonsynonymous	V91I		Thyroid tumor	Benign
200610-21	*CO3*	A9587G	Synonymous	L127L			
2216375	*CO3*	T9682C	Nonsynonymous	M159T			Benign
200610-59	*CO3*	C9785T	Synonymous	Y193Y			
2263321	*CO3*	G9804A	Nonsynonymous	A200T	LHON	Pancreatic cancer cell line	Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2216376	*CO3*	A9855G	Nonsynonymous	I217V			Benign
2216377	*CO3*	T9903C	Nonsynonymous	F233L			Probably damaging
2216378	*CO3*	G9966A	Nonsynonymous	V254I			Benign
2216379	*tRNA-Gly*	T10007C					
2216380	*tRNA-Gly*	T10031C					
2263322	*tRNA-Gly*	T10034C					
2216384	*ND3*	A10086G	Nonsynonymous	N10D	Hypertensive end-stage renal disease		Probably damaging
200610-133	*ND3*	T10118C	Synonymous	I20I			
2216385	*ND3*	G10143A	Nonsynonymous	G29S			Benign
2216386	*ND3*	T10321C	Nonsynonymous	V88A		Bladder tumor	Benign
2216387	*ND3*	T10345C	Nonsynonymous	I96T			Benign
200610-22	*ND3*	A10397G	Synonymous	W113W			
2216389	*tRNA-Arg*	T10410C					
rs28358279	*tRNA-Arg*	T10463C				Endometrium tumor	
MitoA10551G	*ND4L*	A10550G	Synonymous	M27M		Endometrial control tissue	
200610-60	*ND4L*	C10607T	Synonymous	L46L			
2216391	*ND4L*	T10609C	Nonsynonymous	M47T			Benign
200610-135	*ND4L*	T10640C	Synonymous	N57N		Endometrium tumor	
200610-23	*ND4L*	A10754G	Synonymous	L95L			
200610-92	*ND4*	G11016A	Nonsynonymous	S86N		Thyroid tumor	Benign
2216394	*ND4*	T11025C	Nonsynonymous	L89P			Benign
200610-61	*ND4*	C11061T	Nonsynonymous	S101F			Benign
2216396	*ND4*	A11172G	Nonsynonymous	N138S			Probably damaging
2216397	*ND4*	C11177T	Nonsynonymous	P140S			Probably damaging
2216398	*ND4*	T11204C	Nonsynonymous	F149L			Benign
MitoA11252G	*ND4*	A11251G	Synonymous	L164L			
2216399	*ND4*	T11253C	Nonsynonymous	I165T	LHON; PD		Benign
200610-63	*ND4*	C11288T	Synonymous	L177L			
MitoA11468G	*ND4*	A11467G	Synonymous	L236L	Altered brain pH; sCJD patients		

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
200610-65	*ND4*	C11536T	Synonymous	Y259Y			
2010-08-MT-82	*ND4*	A11560G	Synonymous	W267W			
2216400	*ND4*	G11696A	Nonsynonymous	V313I	LHON; LDYT; DEAF; Hypertension helper mutation		Benign
200610-25	*ND4*	A11959G	Synonymous	M400M			
2216401	*ND4*	G11963A	Nonsynonymous	V402I			Benign
2216402	*ND4*	G11969A	Nonsynonymous	A404T			Benign
2216403	*ND4*	T11984C	Nonsynonymous	Y409H			Probably damaging
200610-136	*ND4*	T12121C	Synonymous	I454I			
2216406	*tRNA-Ser*	G12236A			DEAF	Thyroid tumor	
2216407	*tRNA-Ser*	A12248G					
200610-137	*tRNA-Leu2*	T12285C					
rs2853498	*tRNA-Leu2*	A12308G			CPEO; Stroke; CM; Breast, Renal, & Prostate cancer risk; Altered brain pH; sCJD; Haplogroup U marker	Endometrium control tissue; Lung tumor; Prostate tumor	
2216417	*ND5*	G12454A	Nonsynonymous	V40I			Benign
2216418	*ND5*	C12542T	Nonsynonymous	A69V			Possibly damaging
2010-08-MT-143	*ND5*	A12642G	Synonymous	E102E			
200610-94	*ND5*	G12771A	Synonymous	E145E			
2263325	*ND5*	G12820A	Nonsynonymous	A162T			Benign
200610-27	*ND5*	A12822G	Synonymous	A162A			
2263326	*ND5*	A12937G	Nonsynonymous	M201V			Benign
2216420	*ND5*	G12940A	Nonsynonymous	A202T			Benign
2216421	*ND5*	A12950G	Nonsynonymous	N205S			Benign
2216422	*ND5*	A13105G	Nonsynonymous	I257V			Benign
2216423	*ND5*	G13135A	Nonsynonymous	A267T	Possible HCM susceptibility		Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
200610-30	*ND5*	A13183G	Nonsynonymous	I283V			Benign
200610-138	*ND5*	T13215C	Synonymous	L293L			
2216424	*ND5*	A13276G	Nonsynonymous	M314V			Benign
rs28359178	*ND5*	G13708A	Nonsynonymous	A458T	LHON; Increased MS risk; Higher frequency in PD-ADS; Haplogroup J marker	Acute leukemia platelets, leukocytes, & bone marrow; Breast tumor	Benign
2010-08-MT-204	*ND5*	G13759A	Nonsynonymous	A475T			Benign
2263327	*ND5*	A13780G	Nonsynonymous	I482V			Benign
2216429	*ND5*	T13789C	Nonsynonymous	Y485H			Probably damaging
2216431	*ND5*	C13880A	Nonsynonymous	S515Y			Possibly damaging
2216432	*ND5*	T13886C	Nonsynonymous	L517P			Benign
2216433	*ND5*	C13924T	Nonsynonymous	P530S			Probably damaging
2216434	*ND5*	A13933G	Nonsynonymous	T533A			Benign
2263328	*ND5*	C13934T	Nonsynonymous	T533M			Benign
2216436	*ND5*	A13942G	Nonsynonymous	T536A			Benign
2216437	Non-coding	G13958C	Nonsynonymous	G541A			Possibly damaging
2263329	*ND5*	A13966G	Nonsynonymous	T544A			Benign
2216439	*ND5*	C13981T	Nonsynonymous	P549S			Benign
2216440	*ND5*	T14000A	Nonsynonymous	L555Q			Probably damaging
200610-139	*ND5*	T14025C	Synonymous	P563P			
2216441	*ND5*	A14053G	Nonsynonymous	T573A		Prostate tumor	Benign
2216442	*ND5*	A14059G	Nonsynonymous	I575V			Benign
2216443	*ND5*	A14148G	Premature Termination	Term604Term			
2216444	*ND6*	T14178C	Nonsynonymous	I166V			Benign
2010-08-MT-232	*ND6*	T14182C	Synonymous	V164V			
2216446	*ND6*	G14258A	Nonsynonymous	P139L			Benign
2216447	*ND6*	T14318C	Nonsynonymous	N119S			Benign
200610-96	*ND6*	G14323A	Synonymous	N117N			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
200610-71	*ND6*	C14338T	Synonymous	V112V			
2010-08-MT-251	*ND6*	A14417G	Nonsynonymous	V86A		Papillary thyroid carcinoma	Benign
2216448	*ND6*	T14502C	Nonsynonymous	I58V	LHON		Benign
200610-141	*ND6*	T14512C	Synonymous	M54M			
200610-97	*ND6*	G14544A	Synonymous	L44L			
200610-98	*ND6*	G14569A	Synonymous	S35S			
2263330	*ND6*	A14687G			Mitochondrial myopathy with respiratory failure	Mitochondrial myopathy with respiratory failure	
2263331	*CYB*	A14793G	Nonsynonymous	H16R			Benign
rs28357681	*CYB*	T14798C	Nonsynonymous	F18L		Glioblastoma	Benign
2216456	*CYB*	G14861A	Nonsynonymous	A39T			Benign
2010-08-MT-288	*CYB*	C14872T	Synonymous	I42I			
200610-33	*CYB*	A14890G	Synonymous	G48G			
200610-34	*CYB*	A14893G	Synonymous	L49L			
2216457	*CYB*	A14927G	Nonsynonymous	T61A			Benign
2216458	*CYB*	A14978G	Nonsynonymous	I78V			Benign
2216459	*CYB*	T14979C	Nonsynonymous	I78T			Benign
2216460	*CYB*	A15038G	Nonsynonymous	I98V			Possibly damaging
2216461	*CYB*	T15071C	Nonsynonymous	Y109H			Benign
2216462	*CYB*	T15074C	Nonsynonymous	S110P			Benign
2216463	*CYB*	G15077A	Nonsynonymous	E111K	DEAF		Probably damaging
2263332	*CYB*	G15119A	Nonsynonymous	A125T			Benign
200610-99	*CYB*	G15148A	Synonymous	P134P			
200610-100	*CYB*	G15172A	Synonymous	G142G		Endometrial tumor	
2216465	*CYB*	T15204C	Nonsynonymous	I153T			Benign
rs2853506	*CYB*	A15218G	Nonsynonymous	T158A			Possibly damaging
2216467	*CYB*	A15236G	Nonsynonymous	I164V			Benign
2216468	*CYB*	G15257A	Nonsynonymous	D171N	LHON; Haplogroup J2 marker		Benign
2263333	*CYB*	C15263T	Nonsynonymous	P173S			Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2216470	*CYB*	G15314A	Nonsynonymous	A190T			Benign
2216471	*CYB*	G15323A	Nonsynonymous	A193T			Benign
2216472	*CYB*	A15326G	Nonsynonymous	T194A			Benign
2216473	*CYB*	C15381T	Nonsynonymous	T212I			Benign
2216474	*CYB*	C15402T	Nonsynonymous	T219I			Possibly damaging
2216475	*CYB*	G15431A	Nonsynonymous	A229T			Benign
rs3088309	*CYB*	C15452A	Nonsynonymous	L236I			Benign
2216478	*CYB*	T15479C	Nonsynonymous	F245L			Benign
200610-35	*CYB*	A15487T	Synonymous	P247P			
2216479	*CYB*	G15497A	Nonsynonymous	G251S	EXIT; Obesity		Benign
200610-143	*CYB*	T15514C	Synonymous	Y256Y			
2216481	*CYB*	A15662G	Nonsynonymous	I306V	Complex mitochondriopathy-associated	Breast tumor	Benign
2216482	*CYB*	T15672C	Nonsynonymous	M309T		Breast tumor	Benign
2263334	*CYB*	T15693C	Nonsynonymous	M316T	Possibly LVNC cardiomyopathy-associated	Breast tumor	Benign
rs41337244	*CYB*	A15758G	Nonsynonymous	I338V			Benign
2216485	*CYB*	G15812A	Nonsynonymous	V356M	LHON		Benign
2216486	*CYB*	A15824G	Nonsynonymous	T360A		Breast tumor	Benign
2216487	*CYB*	C15849T	Nonsynonymous	T368I		Breast tumor	Benign
2216489	*CYB*	A15860G	Nonsynonymous	I372V			Benign
2263335	*tRNA-Thr*	T15889C					
2263341	*tRNA-Thr*	C15904T					
2216491	*tRNA-Thr*	T15905C					
2216492	*tRNA-Thr*	A15907G					
rs2853510	*tRNA-Thr*	A15924G			LIMM		
2216494	*tRNA-Thr*	G15927A			MS; DEAF1555 increased penetrance; CHD		

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2263337	*tRNA-Thr*	G15928A			MS; Idiopathic repeat miscarriage; AD protection		
2216495	*tRNA-Thr*	C15939T					
2263336	*tRNA-Thr*	T15940C					
2216496	*tRNA-Thr*	T15941C					
2216497	*tRNA-Thr*	T15942C			Possibly LVNC-associated		
2216498	*tRNA-Thr*	C15946T					
2216499	*tRNA-Thr*	A15951G			LHON modulator		
2216500	*tRNA-Pro*	C15978T					
2216501	*tRNA-Pro*	T16017C					
2010-08-MT-395	Non-coding	C16069T					
2010-08-MT-398	Non-coding	T16086C					
200610-146	Non-coding	T16217C				Prostate tumor	
rs41378955	Non-coding	G16390A			POAG- potential for association	Breast, ovarian tumor	
2010-08-MT-511	Non-coding	A16399G				Gastric carcinoma	
200610-37	Non-coding	A16482G					

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; POLG = Mitochondrial DNA polymerase ɣ; MNGIE = Mitochondrial neurogastrointestinal encephalopathy; LVNC = Left ventricular non-compaction cardiomyopathy; DEAF = Deafness; LHON = Leber Hereditary optic neuropathy; PEO = Progressive external ophthalmoplegia; CPT = Carnitine palmityltransferase deficiency; DMDF = Diabetes mellitus and deafness; HCM = Hypertrophic cardiomyopathy; GDM = Gestational diabetes mellitus; AD = Alzheimer's disease; PD = Parkinson's disease; RP = Retinitis pigmentosa; AMD = Age-related macular degeneration; POAG = Primary open angle glaucoma; LDYT = Leber hereditary optic neuropathy and dystonia; CM = Cardiomyopathy; sCJD = Sporadic Creutzfeldt-Jaokob disease; MS = Multiple sclerosis; PD-ADS = Acquired demyelinating syndromes; HCM = Hypertrophic cardiomyopathy; EXIT = Exercise intolerance; LIMM = Lethal infantile mitochondrial myopathy; CHD = Coronary heart disease

Table S18. Infinium Human Exome v1.2

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs41531144	Non-coding	T217C					
rs41323649	Non-coding	G228A					
rs41528348	Non-coding	C295T				POLG/MNGIE muscle; Glioblastoma	
rs3901846	Non-coding	G499A				Thyroid & prostate tumors	
rs3928305	*16S rRNA*	G1719A					
rs2853516	*ND1*	G3316A	Nonsynonymous	A4T	Diabetes; LHON; PEO		Benign
rs41456348	*tRNA-Gln*	T4336C			AD; PD; Hearing loss & migraine		
rs3021088	*ND2*	G5460A	Nonsynonymous	A331T	AD; PD		Benign
rs2853825	*COX3*	G9477A	Nonsynonymous	V91I		Thyroid tumor	Benign
rs2853826	*ND3*	A10398G	Nonsynonymous	T114A	PD protective factor; Longevity; Altered cell pH; Metabolic syndrome; Breast cancer risk; Haplogroup IJK marker	Thyroid tumor	Benign
rs28358279	*tRNA-Arg*	T10463C				Endometrium tumor	
rs2853498	*tRNA-Leu2*	A12308G			CPEO; Stroke; CM; Breast, Renal, & Prostate cancer risk; Altered brain pH; sCJD; Haplogroup U marker	Endometrium control tissue; Lung & prostate tumors	
rs28359178	*ND5*	G13708A	Nonsynonymous	A458T	LHON; Increased MS risk; Higher frequency in PD-ADS; Haplogroup J marker	Acute leukemia platelets, leukocytes, & bone marrow; Breast tumor	Benign
rs28357681	*CYB*	T14798C	Nonsynonymous	F18L		Glioblastoma	Benign
rs2853506	*CYB*	A15218G	Nonsynonymous	T158A			Possibly damaging
rs3088309	*CYB*	C15452A	Nonsynonymous	L236I			Benign
rs41337244	*CYB*	A15758G	Nonsynonymous	I338V			Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs2853510	*tRNA-Thr*	A15924G			LIMM		
rs41378955	Non-coding	G16390A			POAG- potential for association	Breast, ovarian tumor	

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; LHON = Leber Hereditary optic neuropathy; PEO = Progressive external ophthalmoplegia; AD = Alzheimer's disease; PD = Parkinson's disease; POLG = Mitochondrial DNA polymerase ɣ; MNGIE = Mitochondrial neurogastrointestinal encephalopathy; CPEO = Chronic progressive external ophthalmoplegia syndrome; CM = Cardiomyopathy; sCJD = Sporadic Creutzfeldt-Jaokob disease; MS = Multiple sclerosis; PD-ADS = Acquired demyelinating syndromes; LIMM = Lethal infantile mitochondrial myopathy; POAG = Primary open angle glaucoma

Table S19. Illumina Human Core 12v1

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2010-08-MT-841	Non-Coding	T72C				Aging brains; POLG/PEO & control muscle; Normal tissues	
2010-08-MT-981	Non-Coding	A93G					
200610-102	Non-Coding	T125C				POLG/PEO & control muscle	
2010-08-MT-544	Non-Coding	T199C				Ovarian carcinoma; POLG/MNGIE muscle	
200610-104	Non-Coding	T212C					
2010-08-MT-550	Non-Coding	A215G				Control skeletal muscle; Esophageal cancer	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT217C	Non-Coding	T217C					
MitoG228A	Non-Coding	G228A					
200610-105	Non-Coding	T236C					
200610-106	Non-Coding	T246C					
MitoG247A	Non-Coding	G247A					
MitoC295T	Non-Coding	C295T				POLG/MNGIE muscle; glioblastoma	
MitoC458T	Non-Coding	C456T				Thyroid tumor	
MitoT479C	Non-Coding	T477C				AD brains; Ovarian tumor	
200610-107	Non-Coding	T482C					
2010-08-MT-830	*12S rRNA*	T711C					
200610-42	*12S rRNA*	C722T					
MitoG752A	*12S rRNA*	A750G			SZ-associated		
2010-08-MT-27	*12S rRNA*	A1041G					
200610-108	*12S rRNA*	T1107C					
200610-109	*12S rRNA*	T1119C					
MitoT1191C	*12S rRNA*	T1189C					
2010-08-MT-526	*16S rRNA*	C1721T				Acute leukemia platelets, leukocytes, & bone marrow	
MitoA1738G	*16S rRNA*	A1736G					
MitoC2485T	*16S rRNA*	T2483C					
200610-2	*16S rRNA*	A2880G					
MitoT2887C	*16S rRNA*	T2885C					
MitoG3012A	*16S rRNA*	G3010A			Cyclic vomiting syndrome with migraine		
MitoA3349G	*ND1*	A3348G	Synonymous				
200610-111	*ND1*	T3396C	Synonymous		NSHL; MIDD		

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
200610-112	*ND1*	T3644C	Nonsynonymous	V113A	BD-associated		Benign
200610-113	*ND1*	T3645C	Synonymous				
MitoG3667A	*ND1*	G3666A	Synonymous				
MitoA3721G	*ND1*	A3720G	Synonymous				
200610-114	*ND1*	T3826C	Synonymous				
MitoG3916A	*ND1*	G3915A	Synonymous				
MitoG3919A	*ND1*	G3918A	Synonymous			Breast tumor	
200610-45	*ND1*	C3921T	Synonymous				
MitoC3971T	*ND1*	C3970T	Synonymous				
MitoC3993T	*ND1*	C3992T	Nonsynonymous	T229M		Thyroid tumor	Benign
200610-115	*ND1*	T4023C	Synonymous				
MitoA4025G	*ND1*	A4024G	Nonsynonymous	T240A			Benign
200610-47	*ND1*	C4025T	Nonsynonymous	T240M			Benign
2010-08-MT-655	*ND1*	G4048A	Nonsynonymous	D248N			Benign
200610-116	*ND1*	T4117C	Synonymous				
2010-08-MT-664	*ND1*	A4164G	Synonymous				
200610-117	*ND1*	T4218C	Synonymous				
200610-48	*tRNA-Gln*	C4335T					
MitoT4337C	*tRNA-Gln*	T4336C			AD; PD; Hearing loss & migraine		
200610-49	*ND2*	C4508T	Synonymous				
MitoG4770A	*ND2*	A4769G	Synonymous				
MitoG4821A	*ND2*	G4820A	Synonymous				
MitoA4825G	*ND2*	A4824G	Nonsynonymous	T119A			Possibly damaging
200610-6	*ND2*	A4833G	Nonsynonymous	T122A	Diabetes helper mutation; AD; PD; Haplogroup G marker		Possibly damaging
MitoA4918G	*ND2*	A4917G	Nonsynonymous	N150D	LHON; Insulin resistance; AMD; NRTI-		Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
					PN; Haplogroup T marker		
MitoT4978C	*ND2*	T4977C	Synonymous	L170L			
MitoT5005C	*ND2*	T5004C	Synonymous	L179L			
MitoG5047A	*ND2*	G5046A	Nonsynonymous	V193I			Benign
200610-119	*ND2*	T5048C	Synonymous	V193V			
200610-120	*ND2*	T5108C	Synonymous	T213T			
MitoC5264T	*ND2*	C5263T	Nonsynonymous	A265V			Benign
200610-7	*ND2*	A5301G	Nonsynonymous	I278V			Benign
MitoA5391G	*ND2*	A5390G	Synonymous	M307M			
MitoT5443C	*ND2*	T5442C	Nonsynonymous	F325L			Benign
MitoA5657G	Non-Coding	A5656G					
MitoG5774A	*tRNA-Cys*	G5773A					
200610-9	*tRNA-Tyr*	A5833G					
MitoA5952G	*CO1*	A5951G	Synonymous	G16G			
2010-08-MT-773	*CO1*	T5999C	Synonymous	A32A		Pancreatic cancer cell line; Glioblastoma	
MitoG6027A	*CO1*	G6026A	Synonymous	L41L			
MitoC6046T	*CO1*	C6045T	Synonymous	L48L			
2010-08-MT-776	*CO1*	A6047G	Synonymous	L48L		Pancreatic cancer cell line; Glioblastoma	
200610-51	*CO1*	C6077T	Synonymous	V58V			
MitoT6222C	*CO1*	T6221C	Synonymous	P106P			
200610-121	*CO1*	T6248C	Synonymous	S115S			
MitoG6261A	*CO1*	G6260A	Synonymous	E119E			
200610-80	*CO1*	G6285A	Nonsynonymous	V128I	Prostate cancer		Benign
200610-123	*CO1*	T6392C	Synonymous	N163N			
200610-81	*CO1*	G6446A	Synonymous	T181T			
MitoT6681C	*CO1*	T6680C	Synonymous	T259T			
200610-82	*CO1*	G6734A	Synonymous	M277M			
MitoT6777C	*CO1*	T6776C	Synonymous	H291H		Breast cystic masses	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoA7056G	*CO1*	A7055G	Synonymous	G384G		MNGIE fibroblasts	
2010-08-MT-831	*CO1*	T7142C	Synonymous	H413H			
MitoT7176C	*CO1*	T7175C	Synonymous	F414F			
MitoC7275T	*CO1*	C7274T	Synonymous	G457G			
MitoG7522A	*tRNA-Asp*	G7521A				Thyroid tumor	
200610-84	*CO2*	G7598A	Nonsynonymous	A5T	Possible LHON helper variant		Benign
2010-08-MT-860	*CO2*	T7684C	Synonymous	L33L			
2010-08-MT-861	*CO2*	G7697A	Nonsynonymous	V38I	Possible HCM susceptibility		Benign
MitoA7769G	*CO2*	A7768G	Synonymous	M61M			
200610-85	*CO2*	G7852A	Synonymous	E89E			
2010-08-MT-874	*CO2*	G7853A	Nonsynonymous	V90I			Benign
200610-86	*CO2*	G7859A	Nonsynonymous	D92N	Progressive Encephalomyopathy		Benign
200610-127	*CO2*	T7861C	Synonymous	D92D			
200610-128	*CO2*	T7870C	Synonymous	L95L			
200610-13	*CO2*	A7972G	Synonymous	E129E			
200610-87	*CO2*	G8020A	Synonymous	P145P			
200610-129	*ATP8*	T8404C	Synonymous	I13I			
200610-53	*ATP8*	C8472T	Nonsynonymous	P36L			Benign
200610-14	*ATP8*	A8502G	Nonsynonymous	N46S			Possibly damaging
MitoG8617T	*ATP8*	G8616T	Nonsynonymous	L30F			Unknown
200610-131	*ATP6*	T8793C	Synonymous	P89P			
200610-55	*ATP6*	C8859T	Synonymous	G111G			
MitoA8870G	*ATP6*	A8869G	Nonsynonymous	M115V			Benign
200610-18	*ATP6*	A8946G	Synonymous	M140M			
200610-56	*ATP6*	A8964T	Nonsynonymous	M140I			Possibly damaging

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
200610-89	*ATP6*	G9064A	Nonsynonymous	A180T			Benign
MitoA9073G	*ATP6*	A9072G	Synonymous	S182S			
MitoA9094G	*ATP6*	A9093G	Synonymous	T189T			
200610-57	*ATP6*	C9140T	Nonsynonymous	A205V			Probably damaging
2010-08-MT-969	*ATP6*	A9150G	Synonymous	L208L			
MitoG9378A	*CO3*	A9377G	Synonymous	W57W			
200610-58	*CO3*	C9458T	Synonymous	I84I			
MitoC9541T	*CO3*	T9540C	Synonymous	L112L			
200610-21	*CO3*	A9587G	Synonymous	L127L			
MitoA9668G	*CO3*	A9667G	Nonsynonymous	N154S			Benign
MitoT9699C	*CO3*	T9698C	Synonymous	L164L			
MitoT9717C	*CO3*	T9716C	Synonymous	G170G			
200610-59	*CO3*	C9785T	Synonymous	Y193Y			
MitoT9900C	*CO3*	T9899C	Synonymous	H231H			
MitoT9951C	*CO3*	T9950C	Synonymous	V248V			
MitoT10035C	*tRNA-Gly*	T10034C					
200610-133	*ND3*	T10118C	Synonymous	I20I			
MitoT10239C	*ND3*	T10238C	Synonymous	I60I			
MitoG10311A	*ND3*	G10310A	Synonymous	L84L			
200610-22	*ND3*	A10397G	Synonymous	W113W			
MitoA10551G	*ND4L*	A10550G	Synonymous	M27M		Endometrium control tissue	
MitoG10587A	*ND4L*	G10586A	Synonymous	S39S			
MitoG10590A	*ND4L*	G10589A	Synonymous	L40L			
200610-135	*ND4L*	T10640C	Synonymous	N57N		Endometrium tumor	
MitoG10689A	*ND4L*	G10688A	Synonymous	V73V			
200610-23	*ND4L*	A10754G	Synonymous	L95L			
200610-91	*ND4*	G10914A	Nonsynonymous	C52Y			Benign
MitoT10916C	*ND4*	T10915C	Synonymous	C52C			
200610-92	*ND4*	G11016A	Nonsynonymous	S86N		Thyroid tumor	Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
200610-61	*ND4*	C11061T	Nonsynonymous	S101F			Benign
MitoA11252G	*ND4*	A11251G	Synonymous	L164L			
200610-63	*ND4*	C11288T	Synonymous	L177L			
MitoG11378A	*ND4*	G11377A	Synonymous	K206K			
MitoA11468G	*ND4*	A11467G	Synonymous	L236L	Altered brain pH; sCJD		
200610-65	*ND4*	C11536T	Synonymous	Y259Y			
2010-08-MT-82	*ND4*	A11560G	Synonymous	W267W			
MitoT11900C	*ND4*	T11899C	Synonymous	S380S			
MitoG11915A	*ND4*	G11914A	Synonymous	T385T			
200610-136	*ND4*	T12121C	Synonymous	I454I			
200610-137	*tRNA-Leu2*	T12285C					
MitoG12631A	*ND5*	G12630A	Synonymous	W98W			
2010-08-MT-143	*ND5*	A12642G	Synonymous	E102E			
MitoC12670T	*ND5*	C12669T	Synonymous	D111D			
MitoT12706C	*ND5*	C12705T	Synonymous	I123I		Prostate tumor	
200610-94	*ND5*	G12771A	Synonymous	E145E			
2010-08-MT-158	*ND5*	T12811C	Nonsynonymous	Y159H			Benign
200610-27	*ND5*	A12822G	Synonymous	A162A			
MitoG12851A	*ND5*	A12850G	Nonsynonymous	I172V			Possibly damaging
MitoA13106G	*ND5*	A13105G	Nonsynonymous	I257V			Benign
200610-30	*ND5*	A13183G	Nonsynonymous	I283V			Benign
200610-138	*ND5*	T13215C	Synonymous	L293L			
MitoA13264G	*ND5*	A13263G	Synonymous	Q309Q			
MitoC13651T	*ND5*	C13650T	Synonymous	P438P			
2010-08-MT-204	*ND5*	G13759A	Nonsynonymous	A475T			Benign
MitoT13966C	*ND5*	T13965C	Synonymous	L543L			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2010-08-MT-226	*ND5*	A14133G	Synonymous	L599L			
2010-08-MT-232	*ND6*	T14182C	Synonymous	V164V			
MitoA14234G	*ND6*	A14233G	Synonymous	D147D			
200610-96	*ND6*	G14323A	Synonymous	N117N			
200610-71	*ND6*	C14338T	Synonymous	V112V			
2010-08-MT-251	*ND6*	A14417G	Nonsynonymous	V86A			Benign
200610-141	*ND6*	T14512C	Synonymous	M54M			
200610-97	*ND6*	G14544A	Synonymous	L44L			
200610-98	*ND6*	G14569A	Synonymous	S35S			
2010-08-MT-288	*CYB*	C14872T	Synonymous	I42I			
200610-33	*CYB*	A14890G	Synonymous	G48G			
200610-34	*CYB*	A14893G	Synonymous	L49L			
MitoG15044A	*CYB*	G15043A	Synonymous	G99G	MDD-associated		
200610-99	*CYB*	G15148A	Synonymous	P134P			
200610-100	*CYB*	G15172A	Synonymous	G142G		Endometrial tumor	
MitoA15302G	*CYB*	G15301A	Synonymous	L185L		Tumor	
200610-143	*CYB*	T15514C	Synonymous	Y256Y			
MitoT15671C	*CYB*	T15670C	Synonymous	H308H			
MitoT15785C	*CYB*	T15784C	Synonymous	P346P	POAG-potential for association	Pancreatic cancer cell line; Breast tumor	
MitoC15834T	*CYB*	C15833T	Synonymous	L363L			
MitoC15905T	*tRNA-Thr*	C15904T					
MitoA15925G	*tRNA-Thr*	A15924G					
MitoG15929A	*tRNA-Thr*	G15928A			MS; Idiopathic repeat miscarriage; AD protection; Possible helper mutation		

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoG15931A	*tRNA-Thr*	G15930A					
2010-08-MT-395	Non-Coding	C16069T					
2010-08-MT-398	Non-Coding	T16086C					
MitoG16130A	Non-Coding	G16129A			Cyclic vomiting syndrome with migraine		
MitoT16145C	Non-Coding	T16144C					
MitoG16146A	Non-Coding	G16145A					
MitoC16149T	Non-Coding	C16148T				Aging brains	
MitoA16163G	Non-Coding	A16162G					
MitoA16164G	Non-Coding	A16163G					
200610-146	Non-Coding	T16217C				Prostate tumor	
MitoC16272T	Non-Coding	C16270T			Melanoma patients		
2010-08-MT-502	Non-Coding	T16356C				Glioblastoma	
2010-08-MT-504	Non-Coding	T16362C					
MitoG16393A	Non-Coding	G16391A					
2010-08-MT-511	Non-Coding	A16399G				Gastric carcinoma	
200610-37	Non-Coding	A16482G					

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; POLG = Mitochondrial DNA polymerase ɣ; PEO = Progressive external ophthalmoplegia; MNGIE = Mitochondrial neurogastrointestinal encephalopathy; AD = Alzheimer's disease; SZ = Schizophrenia; NSHL = Non-syndromic hearing loss; MIDD = Maternally inherited diabetes and deafness; BD = Bipolar disorder; PD = Parkinson's disease; LHON = Leber Hereditary optic neuropathy; AMD = Age-related macular degeneration; NRTI = Nucleoside reverse transcriptase inhibitor; HCM = Hypertrophic cardiomyopathy; sCJD = Sporadic Creutzfeldt-Jaokob disease; MDD = Major depressive disorder; POAG = Primary open angle glaucoma; MS = Multiple sclerosis

Table S20. Illumina Human1M Duo v3

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT217C	Non-Coding	T217C					
MitoG228A	Non-Coding	G228A					
MitoG247A	Non-Coding	G247A					
MitoC295T	Non-Coding	C295T				POLG/MNGIE muscle; Glioblastoma	
MitoC458T	Non-Coding	C456T				Thyroid tumor	
MitoC464T	Non-Coding	C462T				Thyroid tumor	
MitoT479C	Non-Coding	T477C				AD brains; Ovarian tumor	
MitoT491C	Non-Coding	T489C				Ovarian carcinoma; Prostate tumor	
MitoG752A	*12S rRNA*	A750G					
MitoA829G	*12S rRNA*	A827G			DEAF; Haplogroup B4b'd marker		
MitoC1050T	*12S rRNA*	C1048T					
MitoT1191C	*12S rRNA*	T1189C					
MitoG1440A	*12S rRNA*	A1438G					
MitoG1721A	*16S rRNA*	G1719A					
MitoA1738G	*16S rRNA*	A1736G					
MitoT2160C	*16S rRNA*	T2158C					
MitoC2485T	*16S rRNA*	T2483C					
MitoG2708A	*16S rRNA*	A2706G					
MitoC2791T	*16S rRNA*	C2789T					
MitoT2887C	*16S rRNA*	T2885C					
MitoG3012A	*16S rRNA*	G3010A			Cyclic vomiting syndrome with migraine		
MitoT3198C	*16S rRNA*	T3197C					
MitoA3349G	*ND1*	A3348G	Synonymous	L14L			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT3395C	*ND1*	T3394C	Nonsynonymous	Y30H	LHON/Diabetes/CPT deficiency/High altitude adaptation	Acute leukemia platelets, leukocytes, & bone marrow	Benign
MitoA3481G	*ND1*	A3480G	Synonymous	K58K		Prostate tumor	
MitoC3595T	*ND1*	C3594T	Synonymous	V96V		Thyroid tumor	
MitoG3667A	*ND1*	G3666A	Synonymous	G120G			
MitoA3721G	*ND1*	A3720G	Synonymous	Q138Q			
MitoG3916A	*ND1*	G3915A	Synonymous	G203G			
MitoG3919A	*ND1*	G3918A	Synonymous	E204E		Breast tumor	
MitoC3971T	*ND1*	C3970T	Synonymous	L222L			
MitoC3993T	*ND1*	C3992T	Nonsynonymous	T229M			Benign
MitoA4025G	*ND1*	A4024G	Nonsynonymous	T240A			Benign
MitoT4337C	*tRNA-Gln*	T4336C			AD; PD; Hearing loss & migraine		
MitoT4562C	*ND2*	T4561C	Nonsynonymous	V31A			Benign
MitoG4770A	*ND2*	A4769G	Synonymous	M100M			
MitoG4821A	*ND2*	G4820A	Synonymous	E117E			
MitoA4825G	*ND2*	A4824G	Nonsynonymous	T119A			Possibly damaging
MitoC4884T	*ND2*	C4883T	Synonymous	P138P	Glaucoma		
MitoA4918G	*ND2*	A4917G	Nonsynonymous	N150D	LHON; Insulin resistance; AMD; NRTI-PN; Haplogroup T marker		Benign
MitoT4978C	*ND2*	T4977C	Synonymous	L170L			
MitoT5005C	*ND2*	T5004C	Synonymous	L179L			
MitoG5047A	*ND2*	G5046A	Nonsynonymous	V193I			Benign
MitoC5264T	*ND2*	C5263T	Nonsynonymous	A265V			Benign
MitoA5391G	*ND2*	A5390G	Synonymous	M307M			
MitoT5443C	*ND2*	T5442C	Nonsynonymous	F325L			Benign
MitoG5461A	*ND2*	G5460A	Nonsynonymous	A331T	AD; PD		Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT5496C	*ND2*	T5495C	Synonymous	F342F			
MitoA5657G	*NC*	A5656G					
MitoG5774A	*tRNA-Cys*	G5773A					
MitoA5952G	*CO1*	A5951G	Synonymous	G16G			
MitoG6027A	*CO1*	G6026A	Synonymous	L41L			
MitoC6046T	*CO1*	C6045T	Synonymous	L48L			
MitoT6153C	*CO1*	T6152C	Synonymous	V83V			
MitoT6222C	*CO1*	T6221C	Synonymous	P106P			
MitoG6261A	*CO1*	G6260A	Synonymous	E119E			
MitoT6681C	*CO1*	T6680C	Synonymous	T259T			
MitoG6735A	*CO1*	G6734A	Synonymous	M277M			
MitoA6753G	*CO1*	A6752G	Synonymous	L283L			
MitoT6777C	*CO1*	T6776C	Synonymous	H291H		Breast cystic masses	
MitoA7056G	*CO1*	A7055G	Synonymous	G384G		MNGIE fibroblasts	
MitoT7176C	*CO1*	T7175C	Synonymous	T424T			
MitoC7275T	*CO1*	C7274T	Synonymous	G457G			
MitoG7522A	*tRNA-Asp*	G7521A				Thyroid tumor	
MitoA7769G	*CO2*	A7768G	Synonymous	M61M			
MitoG8270A	*CO2*	G8269A	Synonymous	Term228Term			
MitoT8278C	Non-Coding	T8277C					
MitoG8617T	*ATP6*	G8616T	Nonsynonymous	L30F			Probably damaging
MitoC8656T	*ATP6*	C8655T	Synonymous	I43I			
MitoA8870G	*ATP6*	A8869G	Nonsynonymous	M115V			Benign
MitoA9073G	*ATP6*	A9072G	Synonymous	S182S			
MitoA9094G	*ATP6*	A9093G	Synonymous	T189T			
MitoG9378A	*CO3*	A9377G	Synonymous	W57W			
MitoC9541T	*CO3*	T9540C	Synonymous	L112L		Tumor	
MitoA9668G	*CO3*	A9667G	Nonsynonymous	N154S			Benign
MitoT9699C	*CO3*	T9698C	Synonymous	L164L			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT9717C	*CO3*	T9716C	Synonymous	G170G			
MitoT9900C	*CO3*	T9899C	Synonymous	H231H			
MitoT9951C	*CO3*	T9950C	Synonymous	V248V			
MitoT10035C	*tRNA-Gly*	T10034C					
MitoA10045G	*tRNA-Gly*	A10044G			SIDS		
MitoT10239C	*ND3*	T10238C	Synonymous	I60I			
MitoG10311A	*ND3*	G10310A	Synonymous	L84L			
MitoT10322C	*ND3*	T10321C	Nonsynonymous	V88A			Benign
MitoG10399A	*ND3*	A10398G	Nonsynonymous	T114A	PD protective factor; Longevity; Altered cell pH; Metabolic syndrome; Breast cancer risk; Haplogroup IJK marker	Thyroid tumor	Benign
MitoT10464C	*tRNA-Arg*	T10463C				Endometrium tumor	
MitoA10551G	*ND4L*	A10550G	Synonymous	M27M		Endometrium control tissue	
MitoG10587A	*ND4L*	G10586A	Synonymous	S39S			
MitoG10590A	*ND4L*	G10589A	Synonymous	L40L			
MitoG10689A	*ND4L*	G10688A	Synonymous	V73V			
MitoC10874T	*ND4*	T10873C	Synonymous	P38P			
MitoT10916C	*ND4*	T10915C	Synonymous	C52C			
MitoA11252G	*ND4*	A11251G	Synonymous	L164L			
MitoG11378A	*ND4*	G11377A	Synonymous	K206K			
MitoA11468G	*ND4*	A11467G	Synonymous	L236L	Altered brain pH; sCJD patients		
MitoT11486C	*ND4*	T11485C	Synonymous	G242G			
MitoT11900C	*ND4*	T11899C	Synonymous	S380S			
MitoG11915A	*ND4*	G11914A	Synonymous	T385T			
MitoA12309G	*tRNA-Leu2*	A12308G			CPEO; Stroke; CM; Breast, renal, & prostate cancer Risk; Altered	Endometrium control tissue; Lung & prostate tumor	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
					brain pH; sCJD; Haplogroup U marker		
MitoG12373A	*ND5*	G12372A	Synonymous	L12L	Altered brain pH; sCJD	Prostate tumor	
MitoG12631A	*ND5*	G12630A	Synonymous	W98W			
MitoC12670T	*ND5*	C12669T	Synonymous	D111D			
MitoT12706C	*ND5*	C12705T	Synonymous	I123I		Prostate tumor	
MitoG12851A	*ND5*	A12850G	Nonsynonymous	I172V			Possibly damaging
MitoA13106G	*ND5*	A13105G	Nonsynonymous	I257V			Benign
MitoA13264G	*ND5*	A13263G	Synonymous	Q309Q			
MitoC13651T	*ND5*	C13650T	Synonymous	P438P			
MitoA13781G	*ND5*	A13780G	Nonsynonymous	I482V			Benign
MitoT13790C	*ND5*	T13789C	Nonsynonymous	Y485H			Probably damaging
MitoT13966C	*ND5*	T13965C	Synonymous	L543L			
MitoT14179C	*ND6*	T14178C	Nonsynonymous	I166V			Benign
MitoA14234G	*ND6*	A14233G	Synonymous	D147D			
MitoA14583G	*ND6*	A14582G	Nonsynonymous	V31A			Benign
MitoT14799C	*CYB*	T14798C	Nonsynonymous	F18L		Glioblastoma	Benign
MitoG15044A	*CYB*	G15043A	Synonymous	G99G	MDD-associated		
MitoA15245G	*CYB*	A15244G	Synonymous	G166G			
MitoG15258A	*CYB*	G15257A	Nonsynonymous	D171N	LHON; Haplogroup J2 marker		Benign
MitoA15302G	*CYB*	G15301A	Synonymous	L185L		tumor	
MitoC15536T	*CYB*	C15535T	Synonymous	N263N		Breast tumor	
MitoT15671C	*CYB*	T15670C	Synonymous	H308H		Breast tumor	
MitoA15759G	*CYB*	A15758G	Nonsynonymous	I338V			Benign
MitoT15785C	*CYB*	T15784C	Synonymous	P346P	POAG-potential for association	Pancreatic cancer cell line; Breast tumor	
MitoC15834T	*CYB*	C15833T	Synonymous	L363L			
MitoC15905T	*tRNA-Thr*	C15904T					
MitoA15925G	*tRNA-Thr*	A15924G			LIMM		

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoG15929A	*tRNA-Thr*	G15928A			MS; Idiopathic repeat miscarriage; AD protection		
MitoG15931A	*tRNA-Thr*	G15930A					
MitoG16130A	Non-Coding	G16129A			Cyclic vomiting syndrome with migraine		
MitoT16145C	Non-Coding	T16144C					
MitoG16146A	Non-Coding	G16145A					
MitoC16149T	Non-Coding	C16148T				Aging brains	
MitoA16163G	Non-Coding	A16162G					
MitoA16164G	Non-Coding	A16163G					
MitoC16184A	Non-Coding	A16183G				Prostate tumor	
MitoC16272T	Non-Coding	C16270T			Melanoma patients		
MitoC16329T	Non-Coding	C16327T					
MitoG16392A	Non-Coding	G16390A			POAG-potential for association	Breast & ovarian tumors	
MitoG16393A	Non-Coding	G16391A					

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; POLG = Mitochondrial DNA polymerase ɣ; MNGIE = Mitochondrial neurogastrointestinal encephalopathy; AD = Alzheimer's disease; DEAF = Deafness; LHON = Leber Hereditary optic neuropathy; CPT = Carnitine palmityltransferase deficiency; PD = Parkinson's disease; AMD = Age-related macular degeneration; NRTI = Nucleoside reverse transcriptase inhibitors; SIDS = Sudden infant death syndrome; sCJD = Sporadic Creutzfeldt-Jaokob disease; CPEO = Chronic progressive external ophthalmoplegia syndrome; CM = Cardiomyopathy; MDD = Major depressive disorder; POAG = Primary open angle glaucoma; LIMM = Lethal infantile mitochondrial myopathy

Table S21. Human Omni 2.5-8v1.1

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2010-08-MT-841	Non-Coding	T72C				Aging brains; POLG, PEO & control muscle; normal tissues	
2010-08-MT-981	Non-Coding	A93G					
200610-102	Non-Coding	T125C				POLG, PEO & control muscle	
2010-08-MT-544	Non-Coding	T199C				Ovarian carcinoma	
200610-104	Non-Coding	T212C					
2010-08-MT-550	Non-Coding	A215G				Esophageal cancer	
200610-105	Non-Coding	T236C					
200610-106	Non-Coding	T246C					
200610-107	Non-Coding	T482C					
2010-08-MT-723	Non-Coding	G513A					
2010-08-MT-729	Non-Coding	A523C					
2010-08-MT-830	*12S rRNA*	T711C					
200610-42	*12S rRNA*	C722T					
200610-75	*12S rRNA*	G951A					
2010-08-MT-27	*12S rRNA*	A1041G					
200610-108	*12S rRNA*	T1107C					
200610-109	*12S rRNA*	T1119C					
2010-08-MT-526	*16S rRNA*	C1721T				Acute leukemia platelets, leukocytes, & bone marrow	
200610-1	*16S rRNA*	A2755G			Possibly LVNC-associated		

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
200610-2	*16S rRNA*	A2880G					
200610-110	*16S rRNA*	T3027C					
200610-3	*16S rRNA*	A3203G					
200610-43	*16S rRNA*	C3206T					
200610-44	*ND1*	C3330T	Synonymous	L8L			
200610-4	*ND1*	A3384G	Synonymous	K26K			
200610-111	*ND1*	T3396C	Synonymous	Y30Y	NSHL; MIDD		
200610-112	*ND1*	T3644C	Nonsynonymous	V113A	BD		Benign
200610-113	*ND1*	T3645C	Synonymous	V113V			
200610-76	*ND1*	G3705A	Synonymous	L133L			
200610-114	*ND1*	T3826C	Synonymous	L174L			
200610-45	*ND1*	C3921T	Synonymous	S205S			
200610-46	*ND1*	C3970T	Synonymous	L222L			
200610-115	*ND1*	T4023C	Synonymous	T239T			
200610-47	*ND1*	C4025T	Nonsynonymous	T240M			Benign
2010-08-MT-655	*ND1*	G4048A	Nonsynonymous	D248N			Benign
200610-116	*ND1*	T4117C	Synonymous	L271L			
2010-08-MT-664	*ND1*	A4164G	Synonymous	M286M			
200610-117	*ND1*	T4218C	Synonymous	Y304Y			
200610-48	*tRNA-Gln*	C4335T					
200610-77	*ND2*	G4491A	Nonsynonymous	V8I			Benign
200610-49	*ND2*	C4508T	Synonymous	I13I			
200610-118	*ND2*	T4646C	Synonymous	Y59Y		Glioblastoma	
200610-6	*ND2*	A4833G	Nonsynonymous	T122A	Diabetes helper mutation; AD; PD; Haplogroup G marker		Possibly damaging
200610-119	*ND2*	T5048C	Synonymous	V193V			
200610-120	*ND2*	T5108C	Synonymous	T213T			
200610-7	*ND2*	A5301G	Nonsynonymous	I278V			Benign
200610-9	*tRNA-tyr*	A5833G					

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2010-08-MT-773	*CO1*	T5999C	Synonymous	A32A		Pancreatic cancer cell line; Glioblastoma	
200610-79	*CO1*	G6026A	Synonymous	L41L			
200610-50	*CO1*	C6045T	Synonymous	L48L			
2010-08-MT-776	*CO1*	A6047G	Synonymous	L48L		Pancreatic cancer cell line; Glioblastoma	
200610-51	*CO1*	C6077T	Synonymous	V58V			
200610-121	*CO1*	T6248C	Synonymous	S115S			
200610-80	*CO1*	G6285A	Nonsynonymous	V128I	Prostate Cancer		Benign
200610-122	*CO1*	T6374C	Synonymous	S157S			
200610-123	*CO1*	T6392C	Synonymous	N163N			
200610-81	*CO1*	G6446A	Synonymous	T181T			
200610-124	*CO1*	T6680C	Synonymous	T259T			
200610-82	*CO1*	G6734A	Synonymous	M277M			
200610-10	*CO1*	A6752G	Synonymous	L283L			
200610-83	*CO1*	G6755A	Synonymous	G284G			
200610-125	*CO1*	T6776C	Synonymous	H291H		Breast cystic masses	
2010-08-MT-831	*CO1*	T7142C	Synonymous	H413H			
200610-52	*tRNA-Ser*	C7476T				Thyroid hyperplasia	
200610-12	*tRNA-Asp*	G7521A				Thyroid tumor	
200610-84	*CO2*	G7598A	Nonsynonymous	A5T	Possible LHON helper		Benign
2010-08-MT-860	*CO2*	T7684C	Synonymous	L33L			
2010-08-MT-861	*CO2*	G7697A	Nonsynonymous	V38I	Possible HCM susceptibility		Benign
200610-126	*CO2*	T7759C	Synonymous	A58A			
2010-08-MT-874	*CO2*	G7853A	Nonsynonymous	V90I			Benign
200610-86	*CO2*	G7859A	Nonsynonymous	D92N	Progressive encephalomyopathy		Benign
200610-127	*CO2*	T7861C	Synonymous	D92D			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
200610-128	*CO2*	T7870C	Synonymous	L95L			
200610-13	*CO2*	A7972G	Synonymous	E129E			
200610-87	*CO2*	G8020A	Synonymous	P145P			
200610-129	*ATP8*	T8404C	Synonymous	I13I			
200610-53	*ATP8*	C8472T	Nonsynonymous	P36L			Benign
200610-14	*ATP8*	A8502G	Nonsynonymous	N46S			Possibly damaging
200610-130	*ATP6*	T8594C	Nonsynonymous	I23T			Possibly damaging
200610-131	*ATP6*	T8793C	Synonymous	P89P			
200610-55	*ATP6*	C8859T	Synonymous	G111G			
200610-18	*ATP6*	A8946G	Synonymous	M140M			
200610-56	*ATP6*	C8964T	Synonymous	T146T			
200610-89	*ATP6*	G9064A	Nonsynonymous	A180T			Benign
200610-57	*ATP6*	C9140T	Nonsynonymous	A205V			Probably damaging
2010-08-MT-969	*ATP6*	A9150G	Synonymous	L208L			
200610-58	*CO3*	C9458T	Synonymous	I84I			
200610-21	*CO3*	A9587G	Synonymous	L127L			
200610-59	*CO3*	C9785T	Synonymous	Y193Y			
200610-133	*ND3*	T10118C	Synonymous	I20I			
200610-22	*ND3*	A10397G	Synonymous	W113W			
200610-60	*ND4L*	C10607T	Synonymous	L46L			
200610-135	*ND4L*	T10640C	Synonymous	N57N		Endometrial tumor	
200610-23	*ND4L*	A10754G	Synonymous	L95L			
200610-92	*ND4*	G11016A	Nonsynonymous	S86N		Thyroid tumor	Benign
200610-61	*ND4*	C11061T	Nonsynonymous	S101F			Benign
200610-62	*ND4*	C11215T	Synonymous	Y152Y			
200610-63	*ND4*	C11288T	Synonymous	L177L			
200610-65	*ND4*	C11536T	Synonymous	Y259Y			
2010-08-MT-82	*ND4*	A11560G	Synonymous	W267W			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
200610-93	*ND4*	G11696A	Nonsynonymous	V313I	LHON; LDYT; DEAF; Hypertension helper mutation		Benign
200610-25	*ND4*	A11959G	Synonymous	M400M			
200610-136	*ND4*	T12121C	Synonymous	I454I			
200610-137	*tRNA-Leu2*	T12285C					
200610-67	*ND5*	C12498T	Synonymous	F54F			
2010-08-MT-143	*ND5*	A12642G	Synonymous	E102E			
200610-94	*ND5*	G12771A	Synonymous	E145E			
2010-08-MT-158	*ND5*	T12811C	Nonsynonymous	Y159H	Possible LHON factor		Benign
200610-27	*ND5*	A12822G	Synonymous	A162A			
200610-68	*ND5*	C12882T	Synonymous	F182F			
200610-95	*ND5*	G12940A	Nonsynonymous	A202T			Benign
200610-30	*ND5*	A13183G	Nonsynonymous	I283V			Benign
200610-138	*ND5*	T13215C	Synonymous	L293L			
200610-69	*ND5*	C13626T	Synonymous	T430T			
200610-70	*ND5*	C13934T	Nonsynonymous	T533M			Benign
2010-08-MT-226	*ND5*	A14133G	Synonymous	L599L			
2010-08-MT-232	*ND6*	T14182C	Synonymous	V164V			
200610-96	*ND6*	G14323A	Synonymous	N117N			
200610-71	*ND6*	C14338T	Synonymous	V112V			
2010-08-MT-251	*ND6*	A14417G	Nonsynonymous	V86A		Papillary thyroid carcinoma	Benign
200610-72	*ND6*	C14433T	Nonsynonymous	A81T			Benign
200610-140	*ND6*	T14502C	Nonsynonymous	I58V	LHON		Benign
200610-141	*ND6*	T14512C	Synonymous	M54M			
200610-97	*ND6*	G14544A	Synonymous	L44L			
200610-98	*ND6*	G14569A	Synonymous	S35S			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
2010-08-MT-288	*CYB*	C14872T	Synonymous	I42I			
200610-33	*CYB*	A14890G	Synonymous	G48G			
200610-34	*CYB*	A14893G	Synonymous	L49L			
200610-142	*CYB*	T14979C	Nonsynonymous	I78T			Benign
200610-99	*CYB*	G15148A	Synonymous	P134P			
200610-100	*CYB*	G15172A	Synonymous	G142G		Endometrial tumor	
200610-74	*CYB*	C15452T	Nonsynonymous	L236F			Benign
200610-35	*CYB*	A15487T	Synonymous	P247P			
200610-143	*CYB*	T15514C	Synonymous	Y256Y			
2010-08-MT-395	Non-Coding	C16069T					
2010-08-MT-398	Non-Coding	T16086C					
2010-08-MT-418	Non-Coding	C16168T					
200610-146	Non-Coding	T16217C				Prostate tumor	
2010-08-MT-463	Non-Coding	C16261T					
2010-08-MT-480	Non-Coding	C16294T					
2010-08-MT-483	Non-Coding	T16298C				Prostate tumor	
2010-08-MT-489	Non-Coding	A16316G					
2010-08-MT-502	Non-Coding	T16356C				Glioblastoma	
2010-08-MT-504	Non-Coding	T16362C					
2010-08-MT-511	Non-Coding	A16399G				Gastric carcinoma	
200610-37	Non-Coding	A16482G					

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; POLG = Mitochondrial DNA polymerase ɣ; PEO = Progressive external ophthalmoplegia; LVNC = Left ventricular non-compaction cardiomyopathy; NSHL = Non-syndromic hearing loss; MIDD = Maternally inherited diabetes and deafness; BD = Bipolar disorder; AD = Alzheimer's disease; PD = Parkinson's disease; LHON = Leber Hereditary optic neuropathy; HCM = Hypertrophic cardiomyopathy; LDYT = Leber hereditary optic neuropathy and dystonia; DEAF = Deafness

Table S22. Illumina Human 660W- Quad v1

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Variants*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT217C	Non-Coding	T217C					
MitoG228A	Non-Coding	G228A					
MitoG247A	Non-Coding	G247A					
MitoC295T	Non-Coding	C295T				POLG & MNGIE muscle; Glioblastoma	
MitoC458T	Non-Coding	C456T				Thyroid tumor	
MitoC464T	Non-Coding	C462T				Thyroid tumor	
MitoT479C	Non-Coding	T477C				AD brains; Ovarian & prostate tumors	
MitoT491C	Non-Coding	T489C				Ovarian carcinoma	
MitoG752A	*12S rRNA*	A750G					
MitoA829G	*12S rRNA*	A827G					
MitoC1050T	*12S rRNA*	C1048T					
MitoT1191C	*12S rRNA*	T1189C					
MitoG1440A	*12S rRNA*	A1438G					
MitoG1721A	*16S rRNA*	G1719A					
MitoA1738G	*16S rRNA*	A1736G					
MitoT2160C	*16S rRNA*	T2158C					
MitoC2485T	*16S rRNA*	T2483C					
MitoG2708A	*16S rRNA*	A2706G					

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Variants*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoC2791T	*16S rRNA*	C2789T					
MitoT2887C	*16S rRNA*	T2885C					
MitoG3012A	*16S rRNA*	G3010A			Cyclic Vomiting Syndrome with Migraine; Reported PM		
MitoT3198C	*16S rRNA*	T3197C					
MitoA3349G	*ND1*	A3348G	Synonymous				
MitoT3395C	*ND1*	T3394C	Nonsynonymous	Y30H			Benign
MitoA3481G	*ND1*	A3480G	Synonymous			Prostate tumor	
MitoC3595T	*ND1*	C3594T	Synonymous		Thyroid tumor		
MitoG3667A	*ND1*	G3666A	Synonymous				
MitoA3721G	*ND1*	A3720G	Synonymous				
MitoG3916A	*ND1*	G3915A	Synonymous				
MitoG3919A	*ND1*	G3918A	Synonymous		Breast tumor		
MitoC3971T	*ND1*	C3970T	Synonymous				
MitoC3993T	*ND1*	C3992T	Nonsynonymous	T229M	Thyroid tumor		Benign
MitoA4025G	*ND1*	A4024G	Nonsynonymous	T240A			Benign
MitoT4337C	*tRNA-Gln*	T4336C			AD; PD; Hearing loss & migraine		
MitoT4562C	*ND2*	T4561C	Nonsynonymous	V31A			Benign
MitoG4770A	*ND2*	A4769G	Synonymous		SZ-associated		
MitoG4821A	*ND2*	G4820A	Synonymous				
MitoA4825G	*ND2*	A4824G	Nonsynonymous	T119A			Possibly damaging
MitoC4884T	*ND2*	C4883T	Synonymous		Glaucoma		
MitoA4918G	*ND2*	A4917G	Nonsynonymous	N150D	LHON; Insulin resistance; AMD; NRTI-PN; Haplogroup T marker		Benign
MitoT4978C	*ND2*	T4977C	Synonymous				
MitoT5005C	*ND2*	T5004C	Synonymous				
MitoG5047A	*ND2*	G5046A	Nonsynonymous	V193I			Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Variants*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoC5264T	*ND2*	C5263T	Nonsynonymous	A265V			Benign
MitoA5391G	*ND2*	A5390G	Synonymous				
MitoT5443C	*ND2*	T5442C	Nonsynonymous	F325L			Benign
MitoG5461A	*ND2*	G5460A	Nonsynonymous	A331T	AD; PD		Benign
MitoT5496C	*ND2*	T5495C	Synonymous				
MitoA5657G	Non-Coding	A5656G					
MitoG5774A	*tRNA-Cys*	G5773A					
MitoA5952G	*CO1*	A5951G	Synonymous				
MitoG6027A	*CO1*	G6026A	Synonymous				
MitoC6046T	*CO1*	C6045T	Synonymous				
MitoT6153C	*CO1*	T6152C	Synonymous				
MitoT6222C	*CO1*	T6221C	Synonymous				
MitoG6261A	*CO1*	G6260A	Synonymous				
MitoT6681C	*CO1*	T6680C	Synonymous				
MitoA6753G	*CO1*	A6752G	Synonymous				
MitoT6777C	*CO1*	T6776C	Synonymous		Breast cystic masses		
MitoA7056G	*CO1*	A7055G	Synonymous		MNGIE fibroblasts		
MitoT7176C	*CO1*	T7175C	Synonymous				
MitoC7275T	*CO1*	C7274T	Synonymous				
MitoG7522A	*tRNA-Asp*	G7521A			Thyroid tumor		
MitoA7769G	*CO2*	A7768G	Synonymous				
MitoG8270A	*CO2*	G8269A	Premature termination	Term228Term			
MitoT8278C	Non-Coding	T8277C					
MitoG8617T	*ATP6*	G8616T	Nonsynonymous	L30F			Probably damaging
MitoC8656T	*ATP6*	C8655T	Synonymous				
MitoA8870G	*ATP6*	A8869G	Nonsynonymous	M115V			Benign
MitoA9073G	*ATP6*	A9072G	Synonymous				
MitoA9094G	*ATP6*	A9093G	Synonymous				
MitoG9378A	*CO3*	A9377G	Synonymous				
MitoC9541T	*CO3*	T9540C	Synonymous		Tumor		
MitoA9668G	*CO3*	A9667G	Nonsynonymous	N154S			Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Variants*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT9699C	*CO3*	T9698C	Synonymous				
MitoT9717C	*CO3*	T9716C	Synonymous				
MitoT9900C	*CO3*	T9899C	Synonymous				
MitoT9951C	*CO3*	T9950C	Synonymous				
MitoT10035C	*tRNA-Gly*	T10034C					
MitoA10045G	*tRNA-Gly*	A10044G			SIDS		
MitoT10239C	*ND3*	T10238C	Synonymous				
MitoG10311A	*ND3*	G10310A	Synonymous				
MitoT10322C	*ND3*	T10321C	Nonsynonymous	V88A	Bladder tumor		Benign
MitoG10399A	*ND3*	A10398G	Nonsynonymous	T114A	PD protective factor; Longevity; Altered cell pH; Metabolic syndrome; Breast cancer risk; Haplogroup IJK marker		Benign
MitoT10464C	*tRNA-Arg*	T10463C			Endometrium tumor		
MitoA10551G	*ND4L*	A10550G	Synonymous		Endometrium control tissue		
MitoG10587A	*ND4L*	G10586A	Synonymous				
MitoG10689A	*ND4L*	G10688A	Synonymous				
MitoT10916C	*ND4*	T10915C	Synonymous				
MitoA11252G	*ND4*	A11251G	Synonymous				
MitoG11378A	*ND4*	G11377A	Synonymous				
MitoA11468G	*ND4*	A11467G	Synonymous		Altered brain pH; sCJD patients		
MitoT11486C	*ND4*	T11485C	Synonymous				
MitoT11900C	*ND4*	T11899C	Synonymous				
MitoG11915A	*ND4*	G11914A	Synonymous				
MitoA12309G	*tRNA-Leu2*	A12308G			CPEO; Stroke; CM; Breast, renal, prostate cancer risk; Altered	Endometrium control tissue; Lung & prostate tumors	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Variants*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoG12373A	*ND5*	G12372A	Synonymous		brain pH; sCJD; Haplogroup U marker Altered brain pH; sCJD patients	Prostate tumor	
MitoG12631A	*ND5*	G12630A	Synonymous				
MitoC12670T	*ND5*	C12669T	Synonymous				
MitoT12706C	*ND5*	C12705T	Synonymous			Prostate tumor	
MitoG12851A	*ND5*	A12850G	Nonsynonymous	I172V			Possibly damaging
MitoA13106G	*ND5*	A13105G	Nonsynonymous	I257V			Benign
MitoA13264G	*ND5*	A13263G	Synonymous				
MitoC13651T	*ND5*	C13650T	Synonymous				
MitoA13781G	*ND5*	A13780G	Nonsynonymous	I482V			Benign
MitoT13790C	*ND5*	T13789C	Nonsynonymous	Y485H			Probably damaging
MitoT13966C	*ND5*	T13965C	Synonymous				
MitoT14179C	*ND6*	T14178C	Nonsynonymous	I166V			Benign
MitoA14234G	*ND6*	A14233G	Synonymous				
MitoA14583G	*ND6*	A14582G	Nonsynonymous	V31A			Benign
MitoT14799C	*CYB*	T14798C	Nonsynonymous	F18L		Glioblastoma	Benign
MitoG15044A	*CYB*	G15043A	Synonymous		MDD-associated		
MitoA15245G	*CYB*	A15244G	Synonymous				
MitoG15258A	*CYB*	G15257A	Nonsynonymous	D171N	LHON; Haplogroup J2 marker; Possible helper mutation		Benign
MitoA15302G	*CYB*	G15301A	Synonymous				
MitoC15536T	*CYB*	C15535T	Synonymous				
MitoT15671C	*CYB*	T15670C	Synonymous			Breast tumor	
MitoA15759G	*CYB*	A15758G	Nonsynonymous	I338V			Benign
MitoT15785C	*CYB*	T15784C	Synonymous		POAG-potential for association	Pancreatic cancer cell line; Breast tumor	
MitoC15834T	*CYB*	C15833T	Synonymous				
MitoC15905T	tRNA-Thr	C15904T					
MitoA15925G	tRNA-Thr	A15924G			LIMM		

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Variants*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoG15929A	tRNA-Thr	G15928A			MS; Idiopathic repeat miscarriage; AD protection; Possible helper mutation		
MitoG15931A	tRNA-Thr	G15930A					
MitoG16130A	Non-Coding	G16129A			Cyclic vomiting syndrome with migraine		
MitoT16145C	Non-Coding	T16144C					
MitoG16146A	Non-Coding	G16145A					
MitoC16149T	Non-Coding	C16148T				Aging brains	
MitoA16163G	Non-Coding	A16162G					
MitoA16164G	Non-Coding	A16163G					
MitoC16184A	Non-Coding	A16183C			Melanoma patients	Lung tumor back-mutation; Prostate tumor	
MitoC16272T	Non-Coding	C16270T			Melanoma patients		
MitoC16329T	Non-Coding	C16327T					
MitoG16392A	Non-Coding	G16390A			POAG-potential for association	Breast & ovarian tumors	
MitoG16393A	Non-Coding	G16391A					

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; POLG = Mitochondrial DNA polymerase ɣ; MNGIE = Mitochondrial neurogastrointestinal encephalopathy; AD = Alzheimer's disease; PD = Parkinson's disease; LHON = Leber Hereditary optic neuropathy; AMD = Age-related macular degeneration; NRTI = Nucleoside reverse transcriptase inhibitor; SZ = Schizophrenia; SIDS = Sudden infant death syndrome; sCJD = Sporadic Creutzfeldt-Jaokob disease; Chronic progressive external ophthalmoplegia; CM = Cardiomyopathy; MDD = Major depressive disorder; POAG = Primary open angle glaucoma; LIMM = Lethal infantile mitochondrial myopathy; MS = Multiple sclerosis

Table S23. Affymetrix 6.0

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs2857291	Non-coding	T195C			BD; Melanoma patients	Elderly fibroblasts; Elderly/AD brains; Lung, thyroid, ovarian, prostate tumors; Glioblastoma	
rs3937037	Non-coding	A235G				Prostate tumor	
rs2853515	Non-coding	A263G				POLG/MNGIE muscle	
rs3883865	Non-coding	T346C					
rs28412942	Non-coding	T408A				Elderly muscle; POLG; PEO; TWINKLE/PEO frontal cortex and muscle; normal tissue	
rs28625645	Non-coding	T489C				Ovarian carcinoma; Prostate tumor	
rs28660704	Non-coding	C497T				Thyroid tumor	
rs28661787	Non-coding	A515G					
rs2853518	*12S rRNA*	A750G					
rs28358570	*12S rRNA*	T921C			Possibly LVNC-associated		
rs28377377	*12S rRNA*	G927A					
rs28579222	*12S rRNA*	A942G					
rs3888511	*12S rRNA*	T961G			Possibly DEAF-associated		
rs28496470	*12S rRNA*	G1273T					
rs28673100	*12S rRNA*	A1374T					
rs28612532	*12S rRNA*	C1378T					
rs28729254	*12S rRNA*	A1379C					
rs28573951	*12S rRNA*	G1380A					
rs28493131	*12S rRNA*	G1389T					
rs28392533	*12S rRNA*	G1410A					
rs28376246	*16S rRNA*	C1699T					

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs2854126	*16S rRNA*	T1700C					
rs28527344	*16S rRNA*	C1703T					
rs28491689	*16S rRNA*	G1750A					
rs28687354	*16S rRNA*	C1833A					
rs28736648	*16S rRNA*	T2277C					
rs28358578	*16S rRNA*	C2332T					
rs28358579	*16S rRNA*	T2352C			Possibly LVNC		
rs28558945	*16S rRNA*	C2415T					
rs28489580	*16S rRNA*	C2516A					
rs28445018	*16S rRNA*	T2630C					
rs2854128	*16S rRNA*	A2706G					
rs28406270	*16S rRNA*	A2753C					
rs28619217	*16S rRNA*	A2755G			Possibly LVNC		
rs2856980	*16S rRNA*	G2758A					
rs28358581	*16S rRNA*	C2789T					
rs3928312	*16S rRNA*	A2833G					
rs2854129	*16S rRNA*	T2857C					
rs28393169	*16S rRNA*	T2863C					
rs28627238	*16S rRNA*	A2866C					
rs2854130	*16S rRNA*	T2885C					
rs28597879	*16S rRNA*	G3001A					
rs28611051	*16S rRNA*	A3125G					
rs28408321	*16S rRNA*	C3126G					
rs28656364	*16S rRNA*	C3149T					
rs28434229	*16S rRNA*	G3196A			AD; PD		
rs2854131	*16S rRNA*	T3197C					
rs28553329	*16S rRNA*	C3210T					
rs28358582	*ND1*	T3308C	Nonsynonymous	M1T	MELAS; DEAF enhancer; Hypertension; LVNC; Putative LHON	Colorectal tumor	Probably damaging
rs28416101	*ND1*	T3336C	Synonymous	I10I			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs28512326	*ND1*	G3436A	Premature termination	TermG44Term		Head & neck tumors	
rs28358583	*ND1*	C3450T	Synonymous	P48P			
rs28358585	*ND1*	A3505G	Nonsynonymous	T67A		Pancreatic cancer cell line; Prostate tumor	Benign
rs28369556	*ND1*	C3513T	Synonymous	T69T			
rs2854133	*ND1*	A3565G	Nonsynonymous	T87A			Benign
rs28647976	*ND1*	C3604G	Nonsynonymous	L100V			Possibly damaging
rs28531858	*ND1*	C3613G	Nonsynonymous	L103V			Probably damaging
rs28357968	*ND1*	G3666A	Synonymous	G120G			
rs28557337	*ND1*	C3696T	Synonymous	I130I			
rs28520658	*ND1*	T3732C	Synonymous	Y142T			
rs28357970	*ND1*	A3796T	Nonsynonymous	T164S			Benign
rs28357972	*ND1*	G3918A	Synonymous	E204E			
rs28429662	*ND2*	C4926T	Nonsynonymous	L153F			Possibly damaging
rs3020561	*ND2*	A4985G	Synonymous	Q172Q		Thyroid tumor	
rs28494478	*ND2*	C5049G	Nonsynonymous	L194V			Probably damaging
rs28456039	*ND2*	A5319G	Nonsynonymous	T284S			
rs28357987	*ND2*	T5393C	Synonymous	S308S			
rs3021088	*ND2*	G5460A	Nonsynonymous	A331T	AD; PD		Benign
rs28588369	*CO1*	G6016A	Nonsynonymous	R38Q			Probably damaging
rs28580752	*CO1*	G6028A	Nonsynonymous	G42D			Probably damaging
rs2856983	*CO1*	G6257A	Synonymous	V118V			
rs28516468	*CO1*	C6455T	Synonymous	F184F			
rs28721398	*CO1*	T6481C	Nonsynonymous	V193A			Benign
rs28461189	*CO1*	C6489G	Nonsynonymous	L196V			Probably damaging
rs28371932	*CO1*	T6505C	Nonsynonymous	V201A			Probably damaging
rs1064597	*ATP6*	C8647G	Nonsynonymous	R41G			Probably damaging
rs28479867	*ATP6*	G8648A	Nonsynonymous	R41Q			Probably damaging
rs2853822	*ATP6*	C8655T	Synonymous	I43I			
rs28624611	*ATP6*	G8719A	Premature Termination	GXTer			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs2001031	*ATP6*	A8860G	Nonsynonymous	T112A			Benign
rs28358269	*ATP6*	A9072G	Synonymous	S182S			
rs28434399	*CO3*	C9314T	Synonymous	H36H			
rs28474779	*CO3*	A9336G	Nonsynonymous	M44V			Benign
rs2856984	*CO3*	C9559G	Nonsynonymous	P118R			Probably damaging
rs2856985	*CO3*	G9755A	Synonymous	E183E			
rs2854139	*CO3*	C9818T	Synonymous	H204H			
rs28411821	*CO3*	T9824C	Synonymous	L206L			
rs28690056	*CO3*	T9909C	Nonsynonymous	F235L			Probably damaging
rs28580363	*CO3*	G9912A	Nonsynonymous	E236K			Probably damaging
rs28715301	*CO3*	G9942A	Nonsynonymous	D246N		Endometrial tumor	Probably damaging
rs3134801	*CO3*	T9950C	Synonymous	V248V			
rs28374827	*tRNA-Gly*	T10015A					
rs28358274	*ND3*	A10086G	Nonsynonymous	N10D	Hypertension end-stage renal disease		Probably damaging
rs3899188	*ND3*	T10115C	Synonymous	I19I			
rs28409867	*ND3*	A10133C	Synonymous	P25P			
rs28754574	*ND3*	A10135G	Nonsynonymous	Q26R			Possibly damaging
rs28358275	*ND3*	T10237C	Nonsynonymous	I60T	LHON		Probably damaging
rs28655588	*ND3*	G10260A	Nonsynonymous	E68K			Possibly damaging
rs28673954	*ND3*	T10370C	Synonymous	Y104Y			
rs28358277	*ND3*	G10373A	Synonymous	E105E			
rs2853826	*ND3*	A10398G	Nonsynonymous	T114A	PD protective factor; Longevity; Altered cell pH; Metabolic syndrome; Breast cancer risk; Haplogroup IJK marker	Thyroid tumor	Benign
rs2853487	*ND4L*	G10589A	Synonymous	L40L			
rs28532736	*ND4L*	T10590G	Nonsynonymous	F41V			Possibly damaging
rs28645634	*ND4L*	A10656G	Nonsynonymous	M63V			Benign
rs2853488	*ND4L*	G10688A	Synonymous	V73V			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs2857285	*ND4*	T10915C	Synonymous	C52C			
rs2857286	*ND4*	C10984G	Synonymous	L75L			
rs2853489	*ND4*	A11172G	Nonsynonymous	N138S			Probably damaging
rs2853490	*ND4*	G11176A	Synonymous	Q139Q			
rs28617734	*ND4*	C11186T	Synonymous	L143L			
rs28358285	*ND4*	T11299C	Synonymous	T180T			
rs28609979	*ND4*	T11365C	Synonymous	A202A	Found in 1 HCM patient		
rs2853493	*ND4*	A11467G	Synonymous	L236L	Altered brain pH; sCJD patients		
rs28371977	*ND4*	G11474A	Nonsynonymous	G239S			Probably damaging
rs28588421	*ND4*	T11547G	Nonsynonymous	V263G			Benign
rs2853495	*ND4*	G11719A	Synonymous	G320G			
rs28396842	*ND4*	G11766T	Nonsynonymous	R336L			Probably damaging
rs28384199	*ND4*	C11777G	Nonsynonymous	R340G			Probably damaging
rs28439211	*ND4*	C11819G	Nonsynonymous	L354V			Probably damaging
rs28550734	*ND4*	C11840T	Synonymous	L361L			
rs28713729	*ND4*	G11843A	Nonsynonymous	A362T			Probably damaging
rs28359169	*ND4*	G11969A	Nonsynonymous	A404T			Benign
rs2853497	*ND4*	G12007A	Synonymous	W416W			
rs28639786	*ND4*	C12053T	Nonsynonymous	R432W			Probably damaging
rs28695839	*ND4*	C12112G	Synonymous	P451P			
rs3134560	*tRNA-His*	G12192A			MICM		
rs28469108	*tRNA-Leu2*	G12275A					
rs28493891	*tRNA-Leu2*	G12300A			3243 suppressor mutant	Lung carcinoma cybrid	
rs28490236	*ND5*	A12340G	Nonsynonymous	T2A			Unknown
rs3134561	*ND5*	A12361G	Nonsynonymous	T9A	Nonalcoholic fatty liver disease		Unknown
rs28709525	*ND5*	C12394T	Nonsynonymous	L20F			Unknown
rs28617389	*ND5*	G12406A	Nonsynonymous	V24I			Benign
rs28608480	*ND5*	T12477C	Synonymous	S47S	Possible HCM susceptibility		
rs28359172	*ND5*	A12612G	Synonymous	V92V			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs28410409	*ND5*	G12684A	Synonymous	Q116Q			
rs28359173	*ND5*	A12693G	Synonymous	K119K			
rs2853500	*ND5*	A12720G	Synonymous	M128M			
rs28719001	*ND5*	T12880G	Nonsynonymous	F182V			Probably damaging
rs28448767	*ND5*	C13029T	Synonymous	P231P			
rs28477492	*ND5*	T13095G	Synonymous	C523V			
rs28604589	*ND5*	A13269G	Synonymous	G211G			
rs2853502	*ND5*	A13276G	Nonsynonymous	M314V			Benign
rs28371809	*ND5*	C13384T	Nonsynonymous	L350F			Probably damaging
rs28359176	*ND5*	A13485G	Synonymous	M383M			
rs28376363	*ND5*	C13492G	Nonsynonymous	L386V			Benign
rs28359177	*ND5*	G13590A	Synonymous	L418L			
rs2854123	*ND5*	C13650T	Synonymous	P438P			
rs2853813	*ND5*	C13702G	Nonsynonymous	R456G			Probably damaging
rs28359178	*ND5*	G13708A	Nonsynonymous	A458T	LHON; Increased MS risk; Higher frequency in PD-ADS; Haplogroup J marker	Acute leukemia platelets, leukocytes, & bone marrow; Breast tumor	Benign
rs28630861	*ND5*	T13740C	Synonymous	I468I			
rs28562381	*ND5*	G13843T	Nonsynonymous	D503Y			Probably damaging
rs28359185	*ND5*	T14000A	Nonsynonymous	L555Q			Probably damaging
rs3900944	*ND5*	C14049T	Synonymous	I571I			
rs2853814	*ND6*	C14272G	Nonsynonymous	L134F			Probably damaging
rs2853815	*ND6*	C14365G	Synonymous	V103V			
rs2853816	*ND6*	C14368G	Nonsynonymous	L102F			Probably damaging
rs3135030	*ND6*	T14470C	Synonymous	G68G			
rs28496897	*ND6*	A14580G	Synonymous	L32L			
rs28357678	*ND6*	C14668T	Synonymous	M2M	MDD		
rs2853504	*CYB*	A14793G	Nonsynonymous	H16R			Benign
rs28357684	*CYB*	G15043C	Synonymous	G99G			
rs28357687	*CYB*	T15204C	Nonsynonymous	I153T			Benign
rs2853506	*CYB*	A15218G	Nonsynonymous	T158A			Possibly damaging

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
rs3134742	*CYB*	C15223T	Synonymous	D159D			
rs28573847	*CYB*	G15301A	Synonymous	L185L		Tumor	
rs2853507	*CYB*	G15317A	Nonsynonymous	A191T			Benign
rs2853508	*CYB*	A15326G	Nonsynonymous	T194A			Benign
rs2853509	*CYB*	G15431A	Nonsynonymous	A229T			Benign
rs28357370	*CYB*	A15487T	Synonymous	P247P			
rs3134743	*CYB*	C15508T	Synonymous	D254D			
rs28357372	*CYB*	A15607G	Synonymous	K287K		Breast tumor	
rs28357373	*CYB*	T15629C	Synonymous	L295L		Breast tumor	
rs3094280	*CYB*	A15662G	Nonsynonymous	I306V	Complex mitochondriopathy-associated	Breast tumor	Benign
rs28357374	*CYB*	T15670C	Synonymous	H308H		Breast tumor	
rs28357375	*CYB*	T15784C	Synonymous	P346P	POAG- Potential for association	Pancreatic cancer cell line; Breast tumor	
rs28357376	*CYB*	A15824G	Nonsynonymous	T360A		Breast tumor	Benign
rs3094281	*CYB*	A15851G	Nonsynonymous	I369V			Benign
rs2853510	*tRNA-Thr*	A15923G					
rs3094282	*tRNA-Thr*	G15927A			MS; DEAF1555 increased penetrance/CHD		
rs28561372	Non-coding	A15954G					
rs2853511	Non-coding	T16093C			Cyclic vomiting syndrome	Breast & prostate tumors; Normal tissues	
rs3134562	Non-coding	T16140C					
rs34100702	Non-coding	C16184T					
rs2853513	Non-coding	C16223T				Tumor	
rs2857289	Non-coding	C16256T					
rs2857290	Non-coding	C16270T			Melanoma patients		
rs34799580	Non-coding	T16311C				Prostate tumor	

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; BD = Bipolar disorder; AD = Alzheimer's disease; LVNC = Left ventricular non-compaction cardiomyopathy; DEAF = Deafness; PD = Parkinson's disease; MELAS = Mitochondrial encephalomyopathy lactic acidosis and stroke-like episodes; LHON = Leber Hereditary optic neuropathy; sCJD = Sporadic Creutzfeldt-Jaokob disease; MICM = Maternally inherited cardiomyopathy; HCM = Hypertrophic cardiomyopathy; MS = Multiple sclerosis; PD-ADS = Acquired demyelinating syndromes; MDD = Major depressive disorder; CHD = Coronary heart disease

Table S24. Illumina HumanHap550v3.0

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Mutation Associated Phenotypes*	PolyPhen-2 Prediction
rs41473347	Non-coding	C182T				POLG/PEO muscle	
rs41323649	Non-coding	C228T					
rs41334645	Non-coding	G247A					
rs41528348	Non-coding	C295T				POLG/MNGIE muscle; Glioblastoma	
rs41356551	Non-coding	C456T				Thyroid tumor	
rs41402146	Non-coding	C462T				Thyroid tumor	
rs2853518	*12S rRNA*	A750G					
rs28358569	*12S rRNA*	A827G			DEAF; B4b'd marker		
rs2856982	*12S rRNA*	C1018T					
rs2000974	*12S rRNA*	C1048T					
rs2001030	*12S rRNA*	A1438G					
rs3928305	*16S rRNA*	G1719A					
rs193303006	*16S rRNA*	A1736G					
rs28358579	*16S rRNA*	T2352C			Possibly LVNC associated		
rs28445203	*16S rRNA*	A2483G					
rs2854128	*16S rRNA*	A2706T					
rs28358581	*16S rRNA*	G2789A					
rs3928306	*16S rRNA*	G3010A			Cyclic vomiting syndrome with migraine		

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Mutation Associated Phenotypes*	PolyPhen-2 Prediction
rs41423746	*ND1*	A3348G	Synonymous	L14L			
rs28358584	*ND1*	A3480G	Synonymous	K58K		Prostate tumor	
rs28358586	*ND1*	A3547G	Nonsynonymous	I81V			Benign
rs2854134	*ND1*	A3593G	Nonsynonymous	V96A			Benign
rs28357968	*ND1*	G3666A	Synonymous	G120G			
rs41355750	*ND1*	T3720C	Synonymous	Q138Q			
rs41524046	*ND1*	C3915T	Synonymous	G203G			
rs28357972	*ND1*	C3918T	Synonymous	E204E		Breast tumor	
rs879051705	*ND1*	C3992T	Nonsynonymous	T229M		Thyroid tumor	Benign
rs41504646	*ND1*	C4024T	Nonsynonymous	T240S			Benign
rs1117205	*ND1*	A4104G	Synonymous	L266L			
rs3021086	*ND2*	A4769G	Synonymous	M100M			
rs28357977	*ND2*	C4820T	Synonymous	E117E			
rs28357980	*ND2*	A4917G	Nonsynonymous	N150D	LHON; Insulin resistance; AMD; NRTI-PN; Haplogroup T marker		Benign
rs28357984	*ND2*	C5178A	Nonsynonymous	L237M	Longevity; Extraversion MI; AMS protection; Blood iron metabolism; Haplogroup D marker		Probably damaging
rs41320049	*ND2*	G5263A	Nonsynonymous	A265V			Benign
rs41333444	*ND2*	A5390G	Synonymous	M307M			
rs3021088	*ND2*	G5460A	Nonsynonymous	A331T	AD; PD		Benign
rs7340122	*CO1*	A5951G	Synonymous	G16G			
rs28516468	*CO1*	C6455T	Synonymous	F184F			
rs41413745	*CO1*	C6734T	Synonymous	M277M			
rs41332953	*CO1*	T6752C	Synonymous	L283L			
rs1978002	*CO1*	A7055G	Synonymous	G384G		MNGIE fibroblasts	
rs41534044	*CO2*	A7768G	Synonymous	M61M			
rs28358883	*CO2*	G8206A	Synonymous	M207M			
rs41427749	*ATP6*	G8616T	Nonsynonymous	L30F			Probably damaging
rs2853822	*ATP6*	C8655T	Synonymous	I43I			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Mutation Associated Phenotypes*	PolyPhen-2 Prediction
rs41513156	*ATP6*	A9093G	Synonymous	T189T			
rs28380140	*CO3*	A9377G	Synonymous	W57W			
rs2248727	*CO3*	T9540C	Synonymous	L112L		Tumor	
rs41482146	*CO3*	A9667G	Nonsynonymous	N154S			Benign
rs41347846	*tRNA-Gly*	T10034C					
rs41467651	*ND3*	G10310A	Synonymous	L84L			
rs2853826	*ND3*	A10398G	Nonsynonymous	T114A	PD protective factor; Longevity; Altered cell pH; Metabolic syndrome; Breast cancer risk; Haplogroup IJK marker	Thyroid tumor	Benign
rs28358281	*ND4L*	G10586A	Synonymous	S39S			
rs2853487	*ND4L*	C10589T	Synonymous	L40L			
rs2853488	*ND4L*	G10688A	Synonymous	V73V			
rs2857284	*ND4*	T10873C	Synonymous	P38P			
rs193302938	*ND4*	C11377T	Synonymous	K206K			
rs28471078	*ND4*	A11722G	Synonymous	L321L			
rs2853496	*ND4*	C11914T	Synonymous	T385T			
rs28359169	*ND4*	C11969T	Nonsynonymous	A404T			Benign
rs2853499	*ND5*	C12372T	Synonymous	L12L	Altered brain pH; sCJD patients	Prostate tumor	
rs41445245	*ND5*	G12630A	Synonymous	W98W			
rs41369547	*ND5*	C12669T	Synonymous	D111D			
rs28705385	*ND5*	T12850C	Nonsynonymous	I172V			Possibly damaging
rs2854123	*ND5*	G13650A	Synonymous	P438P			
rs28357672	*ND6*	A14212G	Synonymous	V154V			
rs527236043	*CYB*	C15043T	Synonymous	G99G	MDD		
rs28357685	*CYB*	G15110A	Nonsynonymous	A122T			Benign
rs41518645	*CYB*	C15257T	Nonsynonymous	D171N	LHON; Haplogroup J2 marker		Benign
rs28357371	*CYB*	G15535A	Synonymous	N263N			
rs41504845	*CYB*	C15833T	Synonymous	L363L			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Mutation Associated Phenotypes*	PolyPhen-2 Prediction
rs35788393	*tRNA-Thr*	G15904A					
rs527236198	*tRNA-Thr*	G15928A			MS; Idiopathic repeat miscarriage; AD protection		
rs41441949	*tRNA-Thr*	G15930A					
rs41534744	Non-coding	C16129T			Cyclic vomiting syndrome with migraine		
rs41419246	Non-coding	G16145A					
rs2854125	Non-coding	C16147T					
rs41479950	Non-coding	A16163G					
rs28671493	Non-coding	A16183C			Melanoma patients	Lung tumor back-mutation; Prostate tumor	
rs2857290	Non-coding	C16270T			Melanoma patients		
rs41458645	Non-coding	G16278A				Ovarian control tissue	
rs41355449	Non-coding	G16327A					
rs41378955	Non-coding	C16390T			POAG- potential for association	Breast & ovarian tumors	
rs869031877	Non-coding	C16391T					

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; POLG = Mitochondrial DNA polymerase ɣ; PEO = Progressive external ophthalmoplegia; MNGIE = Mitochondrial neurogastrointestinal encephalopathy; DEAF = Deafness; LVNC = Left ventricular non-compaction cardiomyopathy; LHON = Leber Hereditary optic neuropathy; AMD = Age-related macular degeneration; NRTI = Nucleoside reverse transcriptase inhibitor; MI = Myocardial infarction; AMS = ; AD = Alzheimer's disease; PD = Parkinson's disease; sCJD = Sporadic Creutzfeldt-Jaokob disease; MDD = Major depressive disorder; MS = Multiple sclerosis; POAG = Primary open angle glaucoma

Table S25. Illumina CardioMetaboChip

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
mt921	*12S rRNA*	T921C			Possibly LVNC-associated		
mt930	*12S rRNA*	G930A					
mt1018	*12S rRNA*	G1018A					
mt1243	*12S rRNA*	T1243C			Pancreatic cancer cell line		
mt1719	*16S rRNA*	G1719A					
mt1736	*16S rRNA*	A1736G					
mt1738	*16S rRNA*	T1738C				Colorectal tumor	
mt1811	*16S rRNA*	A1811G				Head & neck tumors	
mt1888	*16S rRNA*	G1888A					
mt2332	*16S rRNA*	C2332T					
mt2352	*16S rRNA*	T2352C			Possibly LVNC-associated		
mt2416	*16S rRNA*	T2416C					
mt2706	*16S rRNA*	A2706G					
mt2758	*16S rRNA*	G2758A					
mt2768	*16S rRNA*	A2768G					
mt2789	*16S rRNA*	C2789T					
mt2885	*16S rRNA*	T2885C					
mt3243	*tRNA-Leu1*	A3243G				Elderly brain and muscle; Colon tumor; Oncocytoma	
mt3308	*ND1*	T3308C	Nonsynonymous	M1T	MELAS; DEAF enhancer; Hypertension; LVNC; Putative LHON	Colorectal tumor	Probably damaging
mt3348	*ND1*	A3348G	Synonymous	L14L			
mt3394	*ND1*	T3394C	Nonsynonymous	Y30H	LHON; Diabetes; CPT deficiency; High altitude adaptation	Acute leukemia platelets, leukocytes, & bone marrow	
mt3450	*ND1*	C3450T	Synonymous	P48P			
mt3480	*ND1*	A3480G	Synonymous	K58K		Prostate tumor	
mt3516	*ND1*	C3516A	Synonymous	L70L			
mt3594	*ND1*	C3594T	Synonymous	V96V		Thyroid tumor	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
mt3666	*ND1*	G3666A	Synonymous	G120G			
mt3693	*ND1*	G3693A	Synonymous	L129L			
mt3796	*ND1*	A3796T	Nonsynonymous	T164S			Benign
mt3843	*ND1*	A3843G	Synonymous	W179W			
mt3915	*ND1*	G3915A	Synonymous	G203G			
mt3918	*ND1*	G3918A	Synonymous	E204E		Breast tumor	
mt4104	*ND1*	A4104G	Synonymous	L266L			
mt4312	*tRNA-Ile*	C4312T				Thyroid tumor	
mt4336	*tRNA-Gln*	T4336C				AD; PD; Hearing loss & migraine	
mt4580	*ND2*	G4580A	Synonymous	M37M		Pancreatic cancer cell line	
mt4715	*ND2*	A4715G	Synonymous	G82G			
mt4928	*ND2*	T4928C	Synonymous	L153L			
mt4977	*ND2*	T4977C	Synonymous	L170L			
mt5237	*ND2*	G5237A	Synonymous	P256P			
mt5393	*ND2*	T5393C	Synonymous	S308S			
mt6719	*CO1*	T6719C	Synonymous	G272G			
mt7867	*CO2*	C7867T	Synonymous	S94S			
mt8206	*CO2*	G8206A	Synonymous	M207M			
mt8869	*ATP6*	A8869G	Nonsynonymous	M115V			Benign
mt9055	*ATP6*	G9055A	Nonsynonymous	A177T	PD-protective factor		Possibly damaging
mt9072	*ATP6*	A9072G	Synonymous	S182S			
mt9150	*ATP6*	A9150G	Synonymous	L208L			
mt9347	*CO3*	A9347G	Synonymous	L47L			
mt9540	*CO3*	T9540C	Synonymous	L112L			
mt9554	*CO3*	G9554A	Synonymous	W116W			
mt9698	*CO3*	T9698C	Synonymous	L164L			
mt9755	*CO3*	G9755A	Synonymous	E183E			
mt9818	*CO3*	C9818T	Synonymous	H204H			
mt9899	*CO3*	T9899C	Synonymous	H231H			
mt9950	*CO3*	T9950C	Synonymous	V248V			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
mt10034	*tRNA-Gly*	T10034C					
mt10086	*ND3*	A10086G	Nonsynonymous	N10D	Hypertension end-stage renal disease		Probably damaging
mt10115	*ND3*	T10115C	Synonymous	I19I			
mt10238	*ND3*	T10238C	Synonymous	I60I			
mt10321	*ND3*	T10321C	Nonsynonymous	V88A	Bladder tumor		
mt10373	*ND3*	G10373A	Synonymous	E105E			
mt10400	*ND3*	C10400T	Synonymous	T114T			
mt10463	*tRNA-Arg*	T10463C				Endometrium tumor	
mt10550	*ND4L*	A10550G	Synonymous	M27M		Endometrium control tissue	
mt10664	*ND4L*	C10664T	Synonymous	V65V			
mt10819	*ND4*	A10819G	Synonymous	K20K			
mt10915	*ND4*	T10915C	Synonymous	C52C			
mt11177	*ND4*	C11177T	Nonsynonymous	P140S			Probably damaging
mt11299	*ND4*	T11299C	Synonymous	T180T			
mt11377	*ND4*	G11377A	Synonymous	K206K			
mt11485	*ND4*	T11485C	Synonymous	G242G			
mt11641	*ND4*	A11641G	Synonymous	M294M			
mt11674	*ND4*	C11674T	Synonymous	T305T		Pancreatic cancer cell line; Prostate tumor	
mt11812	*ND4*	A11812G	Synonymous	L351L			
mt11899	*ND4*	T11899C	Synonymous	S380S			
mt11914	*ND4*	G11914A	Synonymous	T385T			
mt11969	*ND4*	G11969A	Nonsynonymous	A404T			Benign
mt12007	*ND4*	G12007A	Synonymous	W416W			
mt12236	*tRNA-Ser*	G12236A			DEAF	Thyroid tumor	
mt12308	*tRNA-Leu2*	A12308G			CPEO; Stroke; CM; Renal, breast, & prostate cancer risk; Altered brain pH; sCJD; Haplogroup K & U marker	Endometrium control tissue; Lung & prostate tumors	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
mt12372	*ND5*	G12372A	Synonymous	L12L	Altered brain pH; sCJD patients	Prostate tumor	
mt12414	*ND5*	T12414C	Synonymous	P26P		Pancreatic cancer cell line; Prostate tumor	
mt12519	*ND5*	T12519C	Synonymous	V61V		Bladder tumor	
mt12612	*ND5*	A12612G	Synonymous	V92V			
mt12633	*ND5*	C12633A	Synonymous	S99S			
mt12693	*ND5*	A12693G	Synonymous	K119K			
mt12705	*ND5*	C12705T	Synonymous	I123I		Prostate tumor	
mt12720	*ND5*	A12720G	Synonymous	M128M			
mt12810	*ND5*	A12810G	Synonymous	W158W			
mt13020	*ND5*	T13020C	Synonymous	G228G			
mt13105	*ND5*	A13105G	Nonsynonymous	I257V			Benign
mt13263	*ND5*	A13263G	Synonymous	Q309Q			
mt13485	*ND5*	A13485G	Synonymous	M383M			
mt13590	*ND5*	G13590A	Synonymous	L418L			
mt13650	*ND5*	C13650T	Synonymous	P438P			
mt13708	*ND5*	G13708A	Nonsynonymous	A458T	LHON; Increased MS risk; Higher frequency in PD-ADS; Haplogroup J marker	Acute leukemia platelets, leukocytes, & bone marrow; Breast tumor	Benign
mt13789	*ND5*	T13789C	Nonsynonymous	Y485H			Probably damaging
mt13803	*ND5*	A13803G	Synonymous	T489T			
mt13880	*ND5*	C13880A	Nonsynonymous	S515Y			Possibly damaging
mt13886	*ND5*	T13886C	Nonsynonymous	L517P			Benign
mt13928	*ND5*	G13928C	Nonsynonymous	S531T			Possibly damaging
mt13934	*ND5*	C13934T	Nonsynonymous	T533M			Benign
mt13966	*ND5*	A13966G	Nonsynonymous	T544A			Benign
mt14000	*ND5*	T14000A	Nonsynonymous	L555Q			Probably damaging
mt14088	*ND5*	T14088C	Synonymous	I584I			
mt14148	*ND5*	A14148G	Synonymous	Term604 Term			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
mt14152	*ND6*	A14152G	Synonymous	N174N			
mt14167	*ND6*	C14167T	Synonymous	E169E			
mt14284	*ND6*	C14284T	Synonymous	E130E			
mt14318	*ND6*	T14318C	Nonsynonymous	N119S			Benign
mt14766	*CYB*	C14766T	Nonsynonymous	T7I			Benign
mt14769	*CYB*	A14769G	Nonsynonymous	N8S			Probably damaging
mt14783	*CYB*	T14783C	Synonymous	L13L			
mt14905	*CYB*	G14905A	Synonymous	M53M			
mt14911	*CYB*	C14911T	Synonymous	Y55Y			
mt15043	*CYB*	G15043A	Synonymous	G99G	MDD-associated		
mt15110	*CYB*	G15110A	Nonsynonymous	A122T			Benign
mt15136	*CYB*	C15136T	Synonymous	G130G			
mt15204	*CYB*	T15204C	Nonsynonymous	I153T			Benign
mt15217	*CYB*	G15217A	Synonymous	G157G			
mt15244	*CYB*	A15244G	Synonymous	G166G			
mt15301	*CYB*	G15301A	Synonymous	L185L		Tumor	
mt15311	*CYB*	A15311G	Nonsynonymous	I189V			Benign
mt15431	*CYB*	G15431A	Nonsynonymous	A229T			Benign
mt15452	*CYB*	C15452A	Nonsynonymous	L236I			Benign
mt15535	*CYB*	C15535T	Synonymous	N263N		Breast tumor	
mt15607	*CYB*	A15607G	Synonymous	K287K		Breast tumor	
mt15670	*CYB*	T15670C	Synonymous	H308H		Breast tumor	
mt15824	*CYB*	A15824G	Nonsynonymous	T360A		Breast tumor	Benign
mt15833	*CYB*	C15833T	Synonymous	L363L			
mt15904	*tRNA-Thr*	C15904T					
mt15924	*tRNA-Thr*	A15924G			LIMM		
mt15928	*tRNA-Thr*	G15928A			MS; Idiopathic repeat miscarriage; AD protection		
mt15942	*tRNA-Thr*	T15942C			Possibly LVNC-associated		
mt16189	Non-coding	T16189C			Diabetes; Cardiomyopathy;	Prostate tumor; Normal buccal swab	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
					Endometrial cancer risk; mtDNA copy number; Metabolic syndrome; Melanoma patients		

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; LVNC = Left ventricular non-compaction cardiomyopathy; MELAS = Mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes; DEAF = Deafness; LHON = Leber Hereditary optic neuropathy; CPT = Carnitine palmityltransferase deficiency; AD = Alzheimer's disease; PD = Parkinson's disease; PD-ADS = Acquired demyelinating syndromes; PEO = Progressive external ophthalmoplegia; CM = Cardiomyopathy; sCJD = Sporadic Creutzfeldt-Jaokob disease; MS = Multiple sclerosis; MDD = Major depressive disorder; LIMM = Lethal infantile mitochondrial myopathy

Table S26. Illumina Human 610 Quad v1

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT217C	Non-Coding	T217C					
MitoG228A	Non-Coding	G228A					
MitoG247A	Non-Coding	G247A					
MitoC295T	Non-Coding	C295T				POLG/MNGIE muscle; Glioblastoma	
MitoC458T	Non-Coding	C456T				Thyroid tumor	
MitoC464T	Non-Coding	C462T				Thyroid tumor	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT479C	Non-Coding	T477C				AD brains; Ovarian tumor	
MitoT491C	Non-Coding	T489C				Ovarian carcinoma; Prostate tumor	
MitoG752A	*12S rRNA*	A750G					
MitoA829G	*12S rRNA*	A827G			DEAF; B4b'd marker		
MitoC1050T	*12S rRNA*	C1048T					
MitoT1191C	*12S rRNA*	T1189C					
MitoG1440A	*12S rRNA*	A1438G					
MitoG1721A	*16S rRNA*	G1719A					
MitoA1738G	*16S rRNA*	A1736G					
MitoT2160C	*16S rRNA*	T2158C					
MitoC2485T	*16S rRNA*	T2483C					
MitoG2708A	*16S rRNA*	A2706G					
MitoC2791T	*16S rRNA*	C2789T					
MitoT2887C	*16S rRNA*	T2885C					
MitoG3012A	*16S rRNA*	G3010A			Cyclic vomiting syndrome with migraine		
MitoT3198C	*16S rRNA*	T3197C					
MitoA3349G	*ND1*	A3348G	Synonymous	L14L			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT3395C	*ND1*	T3394C	Nonsynonymous	Y30H	LHON; Diabetes; CPT deficiency; High altitude adaptation	Acute leukemia platelets, leukocytes, & bone marrow	Benign
MitoA3481G	*ND1*	A3480G	Synonymous	K58K		Prostate tumor	
MitoC3595T	*ND1*	C3594T	Synonymous	V96V		Thyroid tumor	
MitoG3667A	*ND1*	G3666A	Synonymous	G120G			
MitoA3721G	*ND1*	A3720G	Synonymous	Q138Q			
MitoG3916A	*ND1*	G3915A	Synonymous	G203G			
MitoG3919A	*ND1*	G3918A	Synonymous	E204E		Breast tumor	
MitoC3971T	*ND1*	C3970T	Synonymous	L222L			
MitoC3993T	*ND1*	C3992T	Nonsynonymous	T229M		Thyroid tumor	Benign
MitoA4025G	*ND1*	A4024G	Nonsynonymous	T240A			Benign
MitoT4337C	*tRNA-Gln*	T4336C			AD; PD; Hearing loss & migraine		
MitoT4562C	*ND2*	T4561C	Nonsynonymous	V31A			Benign
MitoG4770A	*ND2*	A4769G	Synonymous	M100M			
MitoG4821A	*ND2*	G4820A	Synonymous	E117E			
MitoA4825G	*ND2*	A4824G	Nonsynonymous	T119A			Possibly damaging

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoC4884T	*ND2*	C4883T	Synonymous	P138P	Glaucoma		
MitoA4918G	*ND2*	A4917G	Nonsynonymous	N150D	LHON; Insulin resistance; AMD; NRTI-PN; Haplogroup T marker		Benign
MitoT4978C	*ND2*	T4977C	Synonymous	L170L			
MitoT5005C	*ND2*	T5004C	Synonymous	L179L			
MitoG5047A	*ND2*	G5046A	Nonsynonymous	V193I			Benign
MitoC5264T	*ND2*	C5263T	Nonsynonymous	A265V			Benign
MitoA5391G	*ND2*	A5390G	Synonymous	M307M			
MitoT5443C	*ND2*	T5442C	Nonsynonymous	F325L			Benign
MitoG5461A	*ND2*	G5460A	Nonsynonymous	A331T	AD; PD		Benign
MitoT5496C	*ND2*	T5495C	Synonymous	F342F			
MitoA5657G	*Non-Coding*	A5656G					
MitoG5774A	*tRNA-Cys*	G5773A					
MitoA5952G	*CO1*	A5951G	Synonymous	G16G			
MitoC6046T	*CO1*	C6045T	Synonymous	L48L			
MitoG6027A	*CO1*	G6026A	Synonymous	L41L			
MitoT6153C	*CO1*	T6152C	Synonymous	V83V			
MitoT6222C	*CO1*	T6221C	Synonymous	P106P			
MitoG6261A	*CO1*	G6260A	Synonymous	E119E			
MitoT6681C	*CO1*	T6680C	Synonymous	T259T			
MitoG6735A	*CO1*	G6734A	Synonymous	M277M			
MitoA6753G	*CO1*	A6752G	Synonymous	L283L			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT6777C	*CO1*	T6776C	Synonymous	H291H		Breast cystic masses	
MitoA7056G	*CO1*	A7055G	Synonymous	G384G		MNGIE fibroblasts	
MitoT7176C	*CO1*	T7175C	Synonymous	T424T			
MitoC7275T	*CO1*	C7274T	Synonymous	G457G			
MitoG7522A	*tRNA-Asp*	G7521A				Thyroid tumor	
MitoA7769G	*CO2*	A7768G	Synonymous	M61M			
MitoG8270A	*CO2*	G8269A	Synonymous	Term228Term			
MitoT8278C	Non-Coding	T8277C					
MitoG8617T	*ATP6*	G8616T	Nonsynonymous	L30F			Probably damaging
MitoC8656T	*ATP6*	C8655T	Synonymous	I43I			
MitoA8870G	*ATP6*	A8869G	Nonsynonymous	M115V			Benign
MitoA9073G	*ATP6*	A9072G	Synonymous	S182S			
MitoA9094G	*ATP6*	A9093G	Synonymous	T189T			
MitoG9378A	*CO3*	A9377G	Synonymous	W57W			
MitoC9541T	*CO3*	T9540C	Synonymous	L112L		Tumor	
MitoA9668G	*CO3*	A9667G	Nonsynonymous	N154S			Benign
MitoT9699C	*COX3*	T9698C	Synonymous	L164L			
MitoT9717C	*COX3*	T9716C	Synonymous	G170G			
MitoT9900C	*COX3*	T9899C	Synonymous	H231H			
MitoT9951C	*COX3*	T9950C	Synonymous	V248V			
MitoT10035C	*tRNA-Gly*	T10034C					
MitoA10045G	*tRNA-Gly*	A10044G				SIDS	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT10239C	*ND3*	T10238C	Synonymous	I60I			
MitoG10311A	*ND3*	G10310A	Synonymous	L84L			
MitoT10322C	*ND3*	T10321C	Nonsynonymous	V88A		Bladder tumor	Benign
MitoG10399A	*ND3*	A10398G	Nonsynonymous	T114A	PD protective factor; Longevity; Altered cell pH; Metabolic syndrome; Breast cancer risk; Haplogroup IJK marker	Thyroid tumor	Benign
MitoT10464C	*tRNA-Arg*	T10463C				Endometrial tumor	
MitoA10551G	*ND4L*	A10550G	Synonymous	M27M		Endometrial control tissue	
MitoG10587A	*ND4L*	G10586A	Synonymous	S39S			
MitoG10590A	*ND4L*	G10589A	Synonymous	L40L			
MitoG10689A	*ND4L*	G10688A	Synonymous	V73V			
MitoC10874T	*ND4*	T10873C	Synonymous	P38P			
MitoT10916C	*ND4*	T10915C	Synonymous	C52C			
MitoA11252G	*ND4*	A11251G	Synonymous	L164L			
MitoG11378A	*ND4*	G11377A	Synonymous	K206K			
MitoA11468G	*ND4*	A11467G	Synonymous	L236L	Altered brain pH; sCJD patients		
MitoT11486C	*ND4*	T11485C	Synonymous	G242G			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT11900C	*ND4*	T11899C	Synonymous	S380S			
MitoG11915A	*ND4*	G11914A	Synonymous	T385T			
MitoA12309G	*tRNA-Leu2*	A12308G			CPEO; Stroke; CM; Breast, renal, & prostate cancer risk; Altered brain pH; sCJD; Haplogroup K & U marker	Endometrial control tissue; Lung & prostate tumors	
MitoG12373A	*ND5*	G12372A	Synonymous	L12L	Altered brain pH; sCJD patients	Prostate tumor	
MitoG12631A	*ND5*	G12630A	Synonymous	W98W			
MitoC12670T	*ND5*	C12669T	Synonymous	D111D			
MitoT12706C	*ND5*	C12705T	Synonymous	I123I		Prostate tumor	
MitoG12851A	*ND5*	A12850G	Nonsynonymous	I172V			Possibly damaging
MitoA13106G	*ND5*	A13105G	Nonsynonymous	I257V			Benign
MitoA13264G	*ND5*	A13263G	Synonymous	Q309Q			
MitoC13651T	*ND5*	C13650T	Synonymous	P438P			
MitoA13781G	*ND5*	A13780G	Nonsynonymous	I482V			Benign
MitoT13790C	*ND5*	T13789C	Nonsynonymous	Y485H			Probably damaging

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoT13966C	*ND5*	T13965C	Synonymous	L543L			
MitoT14179C	*ND6*	T14178C	Nonsynonymous	I166V			Benign
MitoA14234G	*ND6*	A14233G	Synonymous	D147D			
MitoA14583G	*ND6*	A14582G	Nonsynonymous	V31A			Benign
MitoT14799C	*CYB*	T14798C	Nonsynonymous	F18L		Glioblastoma	Benign
MitoG15044A	*CYB*	G15043A	Synonymous	G99G	MDD-associated		
MitoA15245G	*CYB*	A15244G	Synonymous	G166G			
MitoG15258A	*CYB*	G15257A	Nonsynonymous	D171N	LHON; Haplogroup J2 marker		Benign
MitoA15302G	*CYB*	G15301A	Synonymous	L185L		Tumor	
MitoC15536T	*CYB*	C15535T	Synonymous	N263N			
MitoT15671C	*CYB*	T15670C	Synonymous	H308H		Breast tumor	
MitoA15759G	*CYB*	A15758G	Nonsynonymous	I338V			Benign
MitoT15785C	*CYB*	T15784C	Synonymous	P346P	POAG- potential for association	Pancreatic cancer cell line; Breast tumor	
MitoC15834T	*CYB*	C15833T	Synonymous	L363L			
MitoC15905T	*tRNA-Thr*	C15904T					
MitoA15925G	*tRNA-Thr*	A15924G				LIMM	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
MitoG15929A	*tRNA-Thr*	G15928A			MS; Idiopathic repeat miscarriage; AD protection		
MitoG15931A	*tRNA-Thr*	G15930A					
MitoG16130A	Non-Coding	G16129A			Cyclic vomiting syndrome with migraine		
MitoT16145C	Non-Coding	T16144C					
MitoG16146A	Non-Coding	G16145A					
MitoC16149T	Non-Coding	C16148T				Aging brains	
MitoA16163G	Non-Coding	A16162G					
MitoA16164G	Non-Coding	A16163G					
MitoC16184A	Non-Coding	A16183C			Melanoma patients	Lung tumor back-mutation; Prostate tumor	
MitoC16272T	Non-Coding	C16270T			Melanoma patients		
MitoC16329T	Non-Coding	C16327T					
MitoG16392A	Non-Coding	G16390A			POAG-potential for association	Breast & ovarian tumors	
MitoG16393A	Non-Coding	G16391A					

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; POLG = Mitochondrial DNA polymerase ɣ; MNGIE = Mitochondrial neurogastrointestinal encephalopathy; AD = Alzheimer's disease; DEAF = Deafness; LHON = Leber Hereditary optic neuropathy; CPT = Carnitine palmityltransferase deficiency; PD = Parkinson's disease; AMD = Age-related macular degeneration; NRTI = Nucleoside reverse transcriptase inhibitor; SIDS = Sudden Infant Death Syndrome; sCJD = Sporadic Creutzfeldt-Jakob disease; CPEO = Chronic progressive external ophthalmoplegia; CM = Cardiomyopathy; MDD = Major depressive disorder; POAG = Primary open angle glaucoma; LIMM = Lethal infantile mitochondrial myopathy; MS = Multiple sclerosis;

Table S27. Affymetrix Axiom UKBiobank

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
79504644	Non-coding	A73G				Aging brains; POLG/PEO & control muscle; Buccal cell, thyroid, & prostate tumors	
89025672	Non-coding	G143A					
52321525	Non-coding	C150T			Longevity; Cervical carcinoma; HPV infection risk	Elderly fibroblasts & leukocytes; Lung, thyroid, & prostate tumors	
52321592	Non-coding	G228A					
34461939	Non-coding	A235G				Prostate Tumor	
34461957	Non-coding	A263G				POLG/MNGIE muscle	
92047869	Non-coding	C295T				POLG/MNGIE muscle; Glioblastoma	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
8902575 6	Non-coding	C462T				Thyroid tumor	
7938165 5	Non-coding	T489C				Ovarian carcinoma; Prostate tumor	
8902567 4	Non-coding	C497T				Thyroid tumor	
8902572 5	Non-coding	A547T					
3446219 6	*12S rRNA*	G709A					
7938165 6	*12S rRNA*	A750G					
3446223 0	*12S rRNA*	G769A					
8902577 4	*12S rRNA*	A827G			DEAF; B4b'd marker		
9143959 7	*12S rRNA*	T961G			Possibly DEAF-associated		
8902576 7	*12S rRNA*	T961C			DEAF; Possibly LVNC-associated		
9143959 8	*12S rRNA*	T980C					
8902577 2	*12S rRNA*	G1018A					
3446168 4	*12S rRNA*	T1243C				Pancreatic cancer cell line	
8902573 6	*12S rRNA*	T1391C			Found in 1 HCM patient		
8902569 6	*12S rRNA*	T1406C				Pancreatic cancer cell line	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
34461788	*12S rRNA*	A1438G			SZ-associated		
89025706	*12S rRNA*	A1555G			DEAF		
79381658	*16S rRNA*	G1719A					
79443409	*16S rRNA*	C1721T				Acute leukemia platelets, leukocytes, & bone marrow	
79381659	*16S rRNA*	A1736G					
34461927	*16S rRNA*	T2158C					
89025695	*16S rRNA*	C2218T					
34461942	*16S rRNA*	T2416C					
34461948	*16S rRNA*	T2483C					
34461959	*16S rRNA*	A2706G					
34461963	*16S rRNA*	G2758A					
92047843	*16S rRNA*	C2789T					
34461972	*16S rRNA*	T2885C					
34461976	*16S rRNA*	G3010A			Cyclic vomiting syndrome with migraine		

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
79443412	*16S rRNA*	T3027C					
79381660	*16S rRNA*	T3197C					
89025759	*tRNA-Leu1*	A3243G			MELAS; LS; DMDF; MIDD; SNHL; CPEO; MM: FSGS; ASD; Cardiac & multi-organ dysfunction	Elderly brain & muscle; Colon tumor; Oncocytoma	
89025737	*ND1*	T3308G	Nonsynonymous	M1Term	SIDS		
89025707	*ND1*	G3316A	Nonsynonymous	A4T	Diabetes; LHON; PEO		Benign
34461994	*ND1*	A3348G	Synonymous	L14L			
34461996	*ND1*	T3394C	Nonsynonymous	Y30H	LHON; Diabetes; CPT deficiency; High altitude adaptation	Acute leukemia platelets, leukocytes, & bone marrow	Benign
89025777	*ND1*	A3395G	Nonsynonymous	Y30C	HCM with hearing loss		Benign
89025735	*ND1*	A3397G	Nonsynonymous	M31V	AD; PD; Possibly LVNC-cardiomyopathy associated		Benign
89025688	*ND1*	T3398C	Nonsynonymous	M31T	DMDF + HCM; GDM; Possibly LVNC-cardiomyopathy associated		Benign
89025676	*ND1*	T3423G	Synonymous	V39V			
89025722	*ND1*	G3460A	Nonsynonymous	A52T	LHON		Probably damaging

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
79381661	*ND1*	A3480G	Synonymous	K58K		Prostate Tumor	
89025718	*ND1*	G3531A	Synonymous	P75P			
92047844	*ND1*	G3591A	Synonymous	L95L			
89025741	*ND1*	C3594T	Synonymous	V96V		Thyroid tumor	
79443419	*ND1*	T3645C	Synonymous	V113V			
79381662	*ND1*	G3666A	Synonymous	G120G			
89025710	*ND1*	G3697A	Nonsynonymous	G131S	MELAS; LS; LDYT		Probably damaging
79381663	*ND1*	A3720G	Synonymous	Q138Q			
89025781	*ND1*	G3733A	Nonsynonymous	E143K	LHON		Possibly damaging
89025762	*ND1*	G3736A	Nonsynonymous	V144I	LHON		Benign
89025739	*ND1*	A3796G	Nonsynonymous	T164A	Adult onset dystonia		Benign
92047845	*ND1*	G3834A	Synonymous	L176L			
89025682	*ND1*	T3866C	Nonsynonymous	I187T	LHON & limb claudication		Benign
79381664	*ND1*	G3915A	Synonymous	G203G			
79381665	*ND1*	G3918A	Synonymous	E204E		Breast tumor	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
9204784 6	*ND1*	C3936T	Synonymous	G210G			
8902573 8	*ND1*	G3946A	Nonsynonymou s	E214K	MELAS		Probably damaging
8902572 9	*ND1*	T3949C	Nonsynonymou s	Y215H	MELAS	Thyroid oncocytoma	Probably damaging
7938166 6	*ND1*	C3970T	Synonymous	L222L			
9204784 7	*ND1*	C3990T	Synonymous	Y228Y			
7938166 7	*ND1*	C3992T	Nonsynonymou s	T229M		Thyroid tumor	Benign
3446203 0	*ND1*	A4024G	Nonsynonymou s	T240A			Benign
8902577 0	*ND1*	A4093G	Nonsynonymou s	T263A			Benign
8902572 3	*ND1*	A4104G	Synonymous	L266L			
8902575 5	*ND1*	T4160C	Nonsynonymou s	L285P	LHON		Probably damaging
8902566 8	*ND1*	C4171A	Nonsynonymou s	L289M	LHON		Probably damaging
8902568 4	*ND1*	T4216C	Nonsynonymou s	Y304H	LHON; Insulin resistance; Possible adaptive high altitude variant	Acute leukemia platelets, leukocytes, & bone marrow	Benign
8902571 4	*tRNA-Ile*	A4300G			MICM		
9204784 8	*tRNA-Ile*	A4310G					

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
79381668	*tRNA-Gln*	T4336C			AD; PD; Hearing loss & migraine		
92047849	*ND2*	A4529T	Synonymous	T20T			
34462060	*ND2*	T4561C	Nonsynonymous	V31A			Benign
34462062	*ND2*	G4580A	Synonymous	M37M		Pancreatic cancer cell line	
89025699	*ND2*	T4639C	Nonsynonymous	I57T			Benign
89025744	*ND2*	A4715G	Synonymous	G82G			
34462075	*ND2*	A4769G	Synonymous	M100M			
89025751	*ND2*	A4793G	Synonymous	M108M			
79381669	*ND2*	G4820A	Synonymous	E117E			
79381670	*ND2*	A4824G	Nonsynonymous	T119A			Possibly damaging
34462083	*ND2*	C4883T	Synonymous	P138P	Glaucoma		
79381671	*ND2*	A4917G	Nonsynonymous	N150D	LHON; Insulin resistance; AMD; NRTI-PN; Haplogroup T marker		Benign
34462088	*ND2*	T4977C	Synonymous	L170L			
79381672	*ND2*	T5004C	Synonymous	L179L			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
34462094	*ND2*	G5046A	Nonsynonymous	V193I			Benign
89025778	*ND2*	G5147A	Synonymous	T226T			
89025691	*ND2*	C5178A	Nonsynonymous	L237M	Longevity; Extraversion MI; AMS protection; Blood iron metabolism; Haplogroup D marker		Probably damaging
89025748	*ND2*	G5231A	Synonymous	L254L		Endometrial control tissue	
34462105	*ND2*	C5263T	Nonsynonymous	A265V			Benign
92047850	*ND2*	C5360T	Synonymous	I297I			
79381673	*ND2*	A5390G	Synonymous	M307M			
79381674	*ND2*	T5442C	Nonsynonymous	F325L			Benign
79381676	*ND2*	T5495C	Synonymous	F342F			
89025677	*tRNA-Ala*	C5633T				Thyroid tumor	
34462122	Non-coding	A5656G					
79381677	*tRNA-Cys*	G5773A					
89025680	*COX1*	G5913A	Nonsynonymous	D4N	Prostate cancer; Hypertension		Benign
34462135	*COX1*	A5951G	Synonymous	G16G			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
79443437	*COX1*	T5999C	Synonymous	A32A		Pancreatic cancer cell line; Glioblastoma	
79443438	*COX1*	A6047G	Synonymous	L48L		Pancreatic cancer cell line; Glioblastoma	
89025671	*COX1*	T6185C	Synonymous	F94F			
79381678	*COX1*	T6221C	Synonymous	P106P			
92047851	*COX1*	T6253C	Nonsynonymous	M117T	Prostate cancer; Enriched in POAG cohort		Benign
89025742	*COX1*	C6371T	Synonymous	S156S			
92047870	*COX1*	C6386T	Synonymous	A161A			
86887358	*COX1*	C6455T	Synonymous	F184F			
89025708	*COX1*	C6528T	Synonymous	L209L			
89025724	*COX1*	T6671C	Synonymous	H256H			
34462179	*COX1*	G6734A	Synonymous	M277M			
34462180	*COX1*	A6752G	Synonymous	L283L			
34462182	*COX1*	T6776C	Synonymous	H291H		Breast cystic masses	
34462190	*COX1*	C7028T	Synonymous	A375A			
34462191	*COX1*	A7055G	Synonymous	G384G		MNGIE fibroblasts	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
8902566 6	*COX1*	T7094C	Synonymous	F397F			
3446220 2	*COX1*	T7175C	Synonymous	T424T			
9204785 2	*COX1*	C7276T	Nonsynonymou s	S458L			Probably damaging
8902576 4	*COX1*	G7444A	Premature Stop Codon	Term514 K	LHON; SNHL; DEAF		
7944344 7	*tRNA-Ser*	C7476T				Thyroid hyperplasia	
9204787 1	*COX2*	T7645C	Synonymous	L20L			
8902577 5	*COX2*	T7657C	Synonymous	H24H			
7938167 9	*COX2*	A7768G	Synonymous	M61M			
8902569 2	*COX2*	C7864T	Synonymous	P93P			
8902576 6	*COX2*	G8269A	Synonymous	Term228Term			
8902570 4	*tRNA-Lys*	A8344G			MERRF; Other-LD; Depressive mood disorder; Leukoencephalopathy; HiCM	Bone marrow; Elderly muscle	
9204785 9	*ATP8*	T8448C	Nonsynonymou s	M28T			Benign
7938168 0	*ATP6*	G8616T	Nonsynonymou s	L30F			Probably damaging
9204786 0	*ATP6*	C8655T	Synonymous	I43I			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
8902568 9	*ATP6*	C8684T	Nonsynonymou s	T53I			Benign
8902573 2	*ATP6*	G8697A	Synonymous	M57M		Thyroid tumor	
3446228 2	*ATP6*	A8869G	Nonsynonymou s	M115V			Benign
8902571 9	*ATP6*	T8993G	Nonsynonymou s	L156R	NARP; Leigh Disease; MILS		Probably damaging
8902574 9	*ATP6*	G8994A	Synonymous	L156L			
3446229 3	*ATP6*	C9042T	Synonymous	H172H			
8902575 7	*ATP6*	G9053A	Nonsynonymou s	S176N			Benign
8902568 5	*ATP6*	G9055A	Nonsynonymou s	A177T	PD protective factor		Possibly damaging
3446229 6	*ATP6*	A9072G	Synonymous	S182S			
8902576 9	*ATP6*	T9090C	Synonymous	S188S			
3446229 9	*ATP6*	A9093G	Synonymous	T189T			
9204786 4	*ATP6*	G9123A	Synonymous	L199L			
8902577 3	*COX3*	A9221G	Synonymous	S5S			
7938168 1	*COX3*	A9377G	Synonymous	W57W			
8902575 8	*COX3*	T9647C	Synonymous	A147A			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
79381683	*COX3*	A9667G	Nonsynonymous	N154S			Benign
79381684	*COX3*	T9698C	Synonymous	L164L			
34462338	*COX3*	T9716C	Synonymous	G170G			
79381685	*COX3*	T9899C	Synonymous	H231H			
79381686	*COX3*	T9950C	Synonymous	V248V			
34461569	*tRNA-Gly*	T10034C					
34461570	*tRNA-Gly*	A10044G			SIDS		
89025697	*ND3*	T10084C	Nonsynonymous	I9T			Benign
92047861	*ND3*	C10142T	Synonymous	N28N			
92047862	*ND3*	A10217G	Synonymous	M53M			
79381687	*ND3*	T10238C	Synonymous	I60I			
79381688	*ND3*	G10310A	Synonymous	L84L			
79381689	*ND3*	T10321C	Nonsynonymous	V88A		Bladder tumor	Benign
92047863	*ND3*	C10394T	Synonymous	D112D			
79381690	*ND3*	A10398G	Nonsynonymous	T114A	PD protective factor; Longevity; Altered cell pH; Metabolic	Thyroid tumor	Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
					syndrome; Breast cancer risk; Haplogroup IJK marker		
8902566 9	*ND3*	C10400T	Synonymous	T114T			
7938169 1	*tRNA-Arg*	T10463C				Endometrial tumor	
3446159 3	*ND4L*	A10550G	Synonymous	M27M		Endometrial control tissue	
3446159 5	*ND4L*	G10586A	Synonymous	S39S			
7938169 2	*ND4L*	G10589A	Synonymous	L40L			
3446160 0	*ND4L*	G10688A	Synonymous	V73V			
8688645 1	*ND4*	T10810C	Synonymous	L17L			
8902567 8	*ND4*	A10819G	Synonymous	K20K			
7938169 3	*ND4*	T10873C	Synonymous	P38P			
7938169 4	*ND4*	T10915C	Synonymous	C52C			
8902576 8	*ND4*	T11025C	Nonsynonymou s	L89P			Benign
3446162 3	*ND4*	A11251G	Synonymous	L164L			
8902575 3	*ND4*	T11299C	Synonymous	T180T			
9204787 3	*ND4*	C11332T	Synonymous	A191A		Thyroid tumor	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
79381695	*ND4*	G11377A	Synonymous	K206K			
79381696	*ND4*	A11467G	Synonymous	L236L	Altered brain pH; sCJD patients		
89025715	*ND4*	C11674T	Synonymous	T305T		Pancreatic cancer cell line; Prostate tumor	
89025779	*ND4*	G11778A	Nonsynonymous	R340H	LHON; Progressive dystonia		Probably damaging
34461648	*ND4*	A11812G	Synonymous	L351L			
79381697	*ND4*	T11899C	Synonymous	S380S			
34461653	*ND4*	G11914A	Synonymous	T385T			
92047855	*ND4*	A11947G	Synonymous	T396T		Prostate tumor	
34461680	*ND5*	G12372A	Synonymous	L12L	Altered brain pH; sCJD patients	Prostate tumor	
89025681	*ND5*	A12397G	Nonsynonymous	T21A	Early onset PD		Unknown
89025686	*ND5*	G12406A	Nonsynonymous	V24I			Benign
86886472	*ND5*	G12501A	Synonymous	M55M			
34461691	*ND5*	A12612G	Synonymous	V92V			
79381699	*ND5*	G12630A	Synonymous	W98W			
89025731	*ND5*	C12633A	Synonymous	S99S			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
34461695	*ND5*	C12669T	Synonymous	D111D			
79381700	*ND5*	C12705T	Synonymous	I123I		Prostate tumor	
89025760	*ND5*	A12810G	Synonymous	W158W			
79381701	*ND5*	A12850G	Nonsynonymous	I172V			Possibly damaging
89025712	*ND5*	T12879C	Synonymous	G181G			
89025701	*ND5*	A12950G	Nonsynonymous	N205S			Benign
92047856	*ND5*	T13020C	Synonymous	G228G			
92047857	*ND5*	A13101C	Synonymous	A255A			
52321475	*ND5*	A13104G	Synonymous	G256G			
34461715	*ND5*	A13105G	Nonsynonymous	I257V			Benign
89025747	*ND5*	A13117G	Nonsynonymous	I261V			Possibly damaging
79381702	*ND5*	A13263G	Synonymous	Q309Q			
89025728	*ND5*	T13500C	Synonymous	G388G		Pancreatic cancer cell line	
89025716	*ND5*	C13506T	Synonymous	Y390Y			
89025711	*ND5*	T13617C	Synonymous	I427I			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
79381703	*ND5*	C13650T	Synonymous	P438P			
34461745	*ND5*	G13708A	Nonsynonymous	A458T	LHON; Increased MS risk; Higher frequency in PD-ADS; Haplogroup J marker	Acute leukemia platelets, leukocytes, & bone marrow; Breast tumor	Benign
79443499	*ND5*	G13759A	Nonsynonymous	A475T			Benign
79381704	*ND5*	A13780G	Nonsynonymous	I482V			Benign
34461750	*ND5*	T13789C	Nonsynonymous	Y485H			Probably damaging
89025700	*ND5*	T13879C	Nonsynonymous	S515P		MNGIE fibroblasts	Benign
79381705	*ND5*	T13965C	Synonymous	L543L			
34461763	*ND5*	A13966G	Nonsynonymous	T544A			Benign
89025754	*ND5*	G14016A	Synonymous	K560K			
92047858	*ND5*	A14070G	Synonymous	S578S			
92047868	*ND5*	T14094C	Synonymous	L586L			
79443502	*ND5*	A14133G	Synonymous	L599L			
92047853	*ND5*	A14139G	Synonymous	L601L			
89025746	*ND6*	C14167T	Synonymous	E169E			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
79381706	*ND6*	T14178C	Nonsynonymous	I166V			Benign
86496743	*ND6*	T14318C	Nonsynonymous	N119S			Benign
92047854	*ND6*	T14470A	Synonymous	G68G			
89025717	*ND6*	T14484C	Nonsynonymous	M64V	LHON		Probably damaging
89025780	*ND6*	T14550C	Nonsynonymous	I42V			Benign
92047872	*ND6*	A14552G	Nonsynonymous	V41A		Pancreatic cancer cell line	Benign
79381708	*ND6*	A14582G	Nonsynonymous	V31A			Benign
92047865	*ND6*	C14620T	Synonymous	G18G		Glioblastoma	
89025675	*tRNA-Glu*	T14674C			Reversible COX deficiency myopathy		
34461803	*CYB*	T14798C	Nonsynonymous	F18L		Glioblastoma	Benign
92047866	*CYB*	G14869A	Synonymous	L41L		Breast tumor	
79443511	*CYB*	C14872T	Synonymous	I42I			
34461806	*CYB*	G14905A	Synonymous	M53M			
79381709	*CYB*	G15043A	Synonymous	G99G	MDD-associated		
92047867	*CYB*	G15148A	Synonymous	P134P			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
8902574 5	*CYB*	A15218G	Nonsynonymou s	T158A			Possibly damaging
8902576 1	*CYB*	A15244G	Synonymous	G166G			
8902572 6	*CYB*	C15250T	Synonymous	Y168Y			
7938171 0	*CYB*	G15257A	Nonsynonymou s	D171N	LHON; Haplogroup J2 marker		Benign
7938171 1	*CYB*	G15301A	Synonymous	L185L		Tumor	
3446182 8	*CYB*	C15452A	Nonsynonymou s	L236I			Benign
9204787 4	*CYB*	T15454C	Synonymous	L236L			
3446183 2	*CYB*	C15535T	Synonymous	N263N			
7938171 2	*CYB*	T15670C	Synonymous	H308H		Breast tumor	
8902569 0	*CYB*	T15693C	Nonsynonymou s	M316T	Possibly LVNC cardiomyopathy-associated	Breast tumor	Benign
3446183 7	*CYB*	A15758G	Nonsynonymou s	I338V			Benign
7938171 3	*CYB*	T15784C	Synonymous	P346P	POAG- potential for association	Pancreatic cancer cell line; Breast tumor	
8902569 8	*CYB*	G15812A	Nonsynonymou s	V356M	LHON		Benign
7938171 4	*CYB*	C15833T	Synonymous	L363L			
8902566 7	*CYB*	G15884A	Nonsynonymou s	A380T			Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
52321556	*tRNA-Thr*	C15904T					
79381715	*tRNA-Thr*	A15924G			LIMM		
79381716	*tRNA-Thr*	G15928A			MS; Idiopathic repeat miscarriage; AD protection		
79381717	*tRNA-Thr*	G15930A					
89025694	*tRNA-Thr*	C15946T					
92047875	Non-coding	A15954C					
89025705	Non-coding	A15954G					
79443519	Non-coding	A16051G					
89025750	Non-coding	A16126C					
79381719	Non-coding	T16144C					
79381720	Non-coding	G16145A					
79381721	Non-coding	C16148T				Aging brains	
34461862	Non-coding	G16153A					
92047876	Non-coding	A16183C			Melanoma patients	Lung tumor back-mutation; Prostate tumor	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Phenotypes*	PolyPhen-2 Prediction
8902570 3	Non-coding	C16193T				Ovarian tumor	
8902572 0	Non-coding	T16243C					
7944352 4	Non-coding	C16261T					
7938172 4	Non-coding	C16270T			Melanoma patients		
8902577 6	Non-coding	T16311C				Prostate tumor	
9204784 2	Non-coding	C16377T					
7944353 1	Non-coding	T16356C				Glioblastoma	
7944353 2	Non-coding	T16362C					
7938172 6	Non-coding	G16391A					
8902572 7	Non-coding	G16526A					

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; POLG = Mitochondrial DNA polymerase ɣ; PEO = Progressive external ophthalmoplegia; HPV = Human papillomavirus; MNGIE = Mitochondrial neurogastrointestinal encephalopathy; DEAF = Deafness; HCM = Hypertrophic cardiomyopathy; SZ = Schizophrenia; MELAS = Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes; LS = Leigh syndrome; DMDF = Diabetes mellitus and deafness; MIDD = Maternally inherited diabetes and deafness; SNHL = Sensorineural hearing loss; CPEO = Chronic progressive external ophthalmoplegia; MM = Mitochondrial myopathy; FSGS = Focal segmental glomerulosclerosis; ASD = Autism spectrum disorder; SIDS = Sudden infant death syndrome; LHON = Leber Hereditary optic neuropathy; CPT = Carnitine palmityltransferase deficiency; AD = Alzheimer's disease; PD = Parkinson's disease; LVNC = Left ventricular non-compaction cardiomyopathy; GDM = Gestational diabetes mellitus;

LDYT = Leber hereditary optic neuropathy and dystonia; MICM = Maternally inherited cardiomyopathy; AMD = Age-related macular degeneration; NRTI = Nucleoside reverse transcriptase inhibitor; MI = Myocardial infarction; POAG = Primary open angle glaucoma; MERRF = Myoclonic epilepsy with ragged red fibers; LD = Leigh disease; HiCM = Histiocytoid cardiomyopathy; NARP = Neuropathy, ataxia, and retinitis pigmentosa; MILS = Maternally inherited Leigh syndrome; sCJD = Sporadic Creutzfeldt-Jaokob disease; MS = Multiple sclerosis; MDD = Major depressive disorder; LIMM = Lethal infantile mitochondrial myopathy

Table S28. Illumina Human 1Mv1

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Variants*	PolyPhen-2 Prediction
13273403	Non-coding	C182T				POLG/PEO muscle	
13273572	Non-coding	T217C					
13273477	Non-coding	G228A					
13273478	Non-coding	G247A					
13273574	Non-coding	T250C					
13273409	Non-coding	C295T				POLG/MNGIE muscle; Glioblastoma	
13273415	Non-coding	C456T				Thyroid tumor	
13273416	Non-coding	C462T				Thyroid tumor	
13273584	Non-coding	T477C				AD brains; Ovarian tumor	
13273585	Non-coding	T489C				Ovarian carcinoma; Prostate tumor	
13273499	*12S rRNA*	A750G					
13273350	*12S rRNA*	A827G			DEAF; B4b'd marker		
13273435	*12S rRNA*	G1018A					
13273359	*12S rRNA*	C1048T					
13273530	*12S rRNA*	T1189C					
13273457	*12S rRNA*	G1438A					
13273473	*16S rRNA*	G1719A					

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Variants*	PolyPhen-2 Prediction
13273322	*16S rRNA*	A1736G					
13273571	*16S rRNA*	T2158C					
13273406	*16S rRNA*	T2352C			Possibly LVNC-associated		
13273407	*16S rRNA*	T2483C					
13273479	*16S rRNA*	A2706G					
13273408	*16S rRNA*	C2789T					
13273575	*16S rRNA*	T2885C					
13273481	*16S rRNA*	G3010A			Cyclic vomiting syndrome with migraine		
13273576	*16S rRNA*	T3197C					
13273328	*ND1*	A3348G	Synonymous	L14L			
13273577	*ND1*	T3394C	Nonsynonymous	Y30H	LHON; Diabetes; CPT deficiency; High altitude adaptation	Acute leukemia platelets, leukocytes, & bone marrow	Benign
13273329	*ND1*	A3480G	Synonymous	K58K		Prostate tumor	
13273331	*ND1*	A3547G	Nonsynonymous	I81V			Benign
13273412	*ND1*	C3594T	Synonymous	V96V		Thyroid tumor	
13273482	*ND1*	G3666A	Synonymous	G120G			
13273332	*ND1*	A3720G	Synonymous	Q138Q			
13273483	*ND1*	G3915A	Synonymous	G203G			
13273484	*ND1*	G3918A	Synonymous	E204E		Breast tumor	
13273413	*ND1*	C3970T	Synonymous	L222L			
13273414	*ND1*	C3992T	Nonsynonymous	T229M		Thyroid tumor	Benign
13273333	*ND1*	A4024G	Nonsynonymous	T240A			Benign
13273334	*ND1*	A4104G	Synonymous	L266L			
13273581	*tRNA-Gln*	T4336C			AD; PD; Hearing loss & migraine		
13273582	*ND2*	T4561C	Nonsynonymous	V31A			Benign
13273583	*ND2*	T4639C	Nonsynonymous	I57T			Benign
13273486	*ND2*	A4769G	Synonymous	M100M			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Variants*	PolyPhen-2 Prediction
13273487	*ND2*	G4820A	Synonymous	E117E			
13273339	*ND2*	A4824G	Nonsynonymous	T119A			Possibly damaging
13273417	*ND2*	C4883T	Synonymous	P138P	Glaucoma		
13273340	*ND2*	A4917G	Nonsynonymous	N150D	LHON; Insulin resistance; AMD; NRTI-PN; Haplogroup T marker		Benign
13273587	*ND2*	T4977C	Synonymous	L170L			
13273588	*ND2*	T5004C	Synonymous	L179L			
13273488	*ND2*	G5046A	Nonsynonymous	V193I			Benign
13273419	*ND2*	C5178A	Nonsynonymous	L237M	Longevity; Extraversion MI; AMS protection; Blood iron metabolism; Haplogroup D marker		Probably damaging
13273420	*ND2*	C5263T	Nonsynonymous	A265V			Benign
13273341	*ND2*	A5390G	Synonymous	M307M			
13273591	*ND2*	T5442C	Nonsynonymous	F325L			Benign
13273491	*ND2*	G5460A	Nonsynonymous	A331T	AD; PD		Benign
13273593	*ND2*	T5495C	Synonymous	F342F			
13273342	Non-coding	A5656G					
13273492	*tRNA-Cys*	G5773A					
13273343	*CO1*	A5951G	Synonymous	G16G			
13273493	*CO1*	G6026A	Synonymous	L41L			
13273422	*CO1*	C6045T	Synonymous	L48L			
13273594	*CO1*	T6071C	Synonymous	V56V			
13273595	*CO1*	T6152C	Synonymous	V83V			
13273596	*CO1*	T6221C	Synonymous	P106P			
13273494	*CO1*	G6260A	Synonymous	E119E			
13273597	*CO1*	T6365C	Synonymous	G154G			
13273424	*CO1*	C6455T	Synonymous	F184F			
13273599	*CO1*	T6680C	Synonymous	T259T			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Variants*	PolyPhen-2 Prediction
13273600	*CO1*	T6719C	Synonymous	G272G			
13273495	*CO1*	G6734A	Synonymous	M277M			
13273345	*CO1*	A6752G	Synonymous	L283L			
13273601	*CO1*	T6776C	Synonymous	H291H		Breast cystic masses	
13273346	*CO1*	A7055G	Synonymous	G384G		MNGIE fibroblasts	
13273604	*CO1*	T7175C	Synonymous	T424T			
13273427	*CO1*	C7256T	Synonymous	N451N			
13273428	*CO1*	C7274T	Synonymous	G457G			
13273498	*tRNA-Asp*	G7521A				Thyroid tumor	
13273348	*CO2*	A7768G	Synonymous	M61M			
13273502	*CO2*	G8206A	Synonymous	M207M			
13273504	*CO2*	G8269A	Synonymous	Term228Term			
13273607	Non-coding	T8277C					
13273507	*ATP6*	G8616T	Nonsynonymous	L30F			Probably damaging
13273432	*ATP6*	C8655T	Synonymous	I43I			
13273351	*ATP6*	A8869G	Nonsynonymous	M115V			Benign
13273352	*ATP6*	A9072G	Synonymous	S182S			
13273353	*ATP6*	A9093G	Synonymous	T189T			
13273354	*CO3*	A9221G	Synonymous	S5S			
13273516	*CO3*	A9377G	Synonymous	W57W			
13273434	*CO3*	T9540C	Synonymous	L112L		Tumor	
13273357	*CO3*	A9667G	Nonsynonymous	N154S			Benign
13273611	*CO3*	T9698C	Synonymous	L164L			
13273612	*CO3*	T9716C	Synonymous	G170G			
13273613	*CO3*	T9899C	Synonymous	H231H			
13273614	*CO3*	T9950C	Synonymous	V248V			
13273520	*tRNA-Gly*	T10034C					
13273284	*tRNA-Gly*	A10044G			SIDS		
13273522	*ND3*	T10238C	Synonymous	I60I			

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Variants*	PolyPhen-2 Prediction
13273436	*ND3*	G10310A	Synonymous	L84L			
13273523	*ND3*	T10321C	Nonsynonymous	V88A		Bladder tumor	Benign
13273437	*ND3*	A10398G	Nonsynonymous	T114A	PD protective factor; Longevity; Altered cellular pH; Metabolic syndrome; Breast cancer risk; ADHD; Haplogroup IJK marker	Thyroid tumor	Benign
13273524	*tRNA-Arg*	T10463C				Endometrial tumor	
13273286	*ND4L*	A10550G	Synonymous	M27M		Endometrial control tissue	
13273438	*ND4L*	G10586A	Synonymous	S39S			
13273439	*ND4L*	G10589A	Synonymous	L40L			
13273440	*ND4L*	G10688A	Synonymous	V73V			
13273360	*ND4*	T10873C	Synonymous	P38P			
13273526	*ND4*	T10915C	Synonymous	C52C			
13273288	*ND4*	A11251G	Synonymous	L164L			
13273443	*ND4*	G11377A	Synonymous	K206K			
13273289	*ND4*	A11467G	Synonymous	L236L	Altered brain pH; sCJD patients		
13273528	*ND4*	T11485C	Synonymous	G242G			
13273364	*ND4*	T11722C	Synonymous	L321L			
13273292	*ND4*	A11812G	Synonymous	L351L			
13273529	*ND4*	T11899C	Synonymous	S380S			
13273444	*ND4*	G11914A	Synonymous	T385T			
13273445	*ND4*	G11969A	Nonsynonymous	A404T			Benign
13273294	*tRNA-Leu(CUN)*	A12308G			CPEO; Stroke; CM; Breast, renal, & prostate cancer risk; Altered brain pH; sCJD; Haplogroup K & U marker	Endometrial control tissue; Lung & prostate tumors	
13273447	*ND5*	G12372A	Synonymous	L12L	Altered brain pH; sCJD patients	Prostate tumor	

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Variants*	PolyPhen-2 Prediction
13273532	*ND5*	T12414C	Synonymous	P26P		Pancreatic cancer cell line; Prostate tumor	
13273450	*ND5*	G12630A	Synonymous	W98W			
13273366	*ND5*	C12669T	Synonymous	D111D			
13273534	*ND5*	C12705T	Synonymous	I123I		Prostate tumor	
13273451	*ND5*	A12850G	Nonsynonymous	I172V			Possibly damaging
13273298	*ND5*	A13105G	Nonsynonymous	I257V			Benign
13273299	*ND5*	A13263G	Synonymous	Q309Q			
13273368	*ND5*	C13650T	Synonymous	P438P			
13273301	*ND5*	A13780G	Nonsynonymous	I482V			Benign
13273539	*ND5*	T13789C	Nonsynonymous	Y485H			Probably damaging
13273541	*ND5*	T13965C	Synonymous	L543L			
13273543	*ND6*	T14178C	Nonsynonymous	I166V			Benign
13273371	*ND6*	T14212C	Synonymous	V154V			
13273304	*ND6*	A14233G	Synonymous	D147D			
13273306	*ND6*	A14582G	Nonsynonymous	V31A			Benign
13273549	*CYB*	C14766T	Nonsynonymous	T7I			Benign
13273550	*CYB*	T14783C	Synonymous	L13L			
13273551	*CYB*	T14798C	Nonsynonymous	F18L		Glioblastoma	Benign
13273461	*CYB*	G15043A	Synonymous	G99G	MDD-associated		
13273462	*CYB*	G15110A	Nonsynonymous	A122T			Benign
13273311	*CYB*	A15218G	Nonsynonymous	T158A			Possibly damaging
13273312	*CYB*	A15244G	Synonymous	G166G			
13273463	*CYB*	G15257A	Nonsynonymous	D171N	LHON; Haplogroup J2 marker		Benign
13273313	*CYB*	G15301A	Synonymous	L185L		Tumor	
13273377	*CYB*	C15535T	Synonymous	N263N			
13273555	*CYB*	T15670C	Synonymous	H308H		Breast tumor	
13273316	*CYB*	A15758G	Nonsynonymous	I338V			Benign

ID	Gene	Variant	Type of Variant	Amino Acid	Base Substitution Associated Phenotypes*	Somatic Variant Associated Variants*	PolyPhen-2 Prediction
13273556	*CYB*	T15784C	Synonymous	P346P	POAG- potential for association	Pancreatic cancer cell line; Breast tumor	
13273378	*CYB*	C15833T	Synonymous	L363L			
13273379	*tRNA-Thr*	C15904T					
13273318	*tRNA-Thr*	A15924G			LIMM		
13273465	*tRNA-Thr*	G15928A			Multiple sclerosis; Idiopathic repeat miscarriage; AD protection		
13273466	*tRNA-Thr*	G15930A					
13273467	Non-coding	G16129A			Cyclic vomiting syndrome with migraine		
13273559	Non-coding	T16144C					
13273468	Non-coding	G16145A					
13273382	Non-coding	C16148T				Aging brains	
13273319	Non-coding	A16162G					
13273320	Non-coding	A16163G					
13273384	Non-coding	A16183C			Melanoma patients	Lung tumor back-mutation; Prostate tumor	
13273395	Non-coding	C16270T			Melanoma patients		
13273396	Non-coding	C16278T				Ovarian control tissue	
13273401	Non-coding	C16327T					
13273471	Non-coding	G16390A			POAG- potential for association	Breast & ovarian tumor	
13273472	Non-coding	G16391A					

*MITOMAP

ND = NADH dehydrogenase; CYB = Cytochrome b; CO = Cytochrome c oxidase; ATP = ATP synthase; POLG = Mitochondrial DNA polymerase ɣ; MNGIE = Mitochondrial neurogastrointestinal encephalopathy; AD = Alzheimer's disease; DEAF = Deafness; LVNC = Left ventricular non-compaction cardiomyopathy; LHON = Leber Hereditary optic neuropathy; CPT = Carnitine palmityltransferase deficiency; PD = Parkinson's disease; AMD = Age-related macular degeneration; NRTI = Nucleoside reverse transcriptase inhibitor; MI = Myocardial infarction; SIDS = Sudden infant death

syndrome; ADHD = Attention deficit-hyperactivity disorder; sCJD = Sporadic Creutzfeldt-Jaokob disease; CPEO = Chronic progressive external ophthalmoplegia; CM = Cardiomyopathy; MDD = Major depressive disorder; POAG = Primary open angle glaucoma; LIMM = Lethal infantile mitochondrial myopathy

Table S29. MT-MeSH Terms Extracted from Literature Lab. The number of MT-nDNA candidate genes selected are based on upper (25^{th}) quartile of Literature Lab's LPF statistics. Moreover, later we conditioned each gene to have at least 15 abstracts.

No	MeSH Term	Number of genes in upper quartile
1	Aldehyde dehydrogenase, mitochondrial	112
2	Aspartate aminotransferase, mitochondrial	17
3	Creatine kinase, mitochondrial form	26
4	DNA, mitochondrial	635
5	Electron transport in mitochondria	448
6	Genes, mitochondrial	139
7	Genome, mitochondrial	99
8	Membrane potential, mitochondrial	580
9	Mitochondria	1,990
10	Mitochondria apoptotic	974
11	Mitochondria, heart	306
12	Mitochondria, liver	352
13	Mitochondria, muscle	409
14	Mitochondrial ADP, ATP translocases	135
15	Mitochondrial carnitine palmitoyltransferase	396
16	Mitochondrial degradation	120
17	Mitochondrial diseases	422
18	Mitochondrial dynamics	221
19	Mitochondrial encephalomyopathies	103
20	Mitochondrial gene expression	102
21	Mitochondrial LC fatty-acid beta-oxidation	2
22	Mitochondrial membrane transport proteins	669
23	Mitochondrial membranes	422
24	Mitochondrial myopathies	167
25	Mitochondrial proteins	1,350
26	Mitochondrial proton- translocating ATPases	200
27	Mitochondrial ribosomes	43
28	Mitochondrial size	66
29	Mitochondrial swelling	157
30	Mitochondrial trifunctional protein	33
31	Mitochondrial trifunctional protein, alpha subunit	4
32	Mitochondrial trifunctional protein, beta subunit	7
33	Mitochondrial turnover	242
34	Mitochondrial uncoupling protein 2	62
35	Mitochondrial uncoupling protein 3	16
36	Mitochondrial uncoupling proteins	308

Table S30. Selected 2,283 Candidate Genes (MT-nDNA) that May Contribute to Mitochondria (**Separate Excel Worksheet**)

Table S31. Final list of genes (HUGO approved names) as mtDNA- and possible MT-genes candidates for contributing to mitochondria and corresponding MIM number and location. (**Separate Excel Worksheet**)

Table S32. Predicted Functional Effects of mtDNA Non-Synonymous Mutations Present in Our Arrays. (**Separate DOCX Table**)

3. Supplementary Methods

1. **Phenotype Preparation**

Seven variables were studied: 1). BMI-body mass index (kg/m^2); 2). WHR- waist to hip ratio, a unit-less measure with the condition that both waist and hip were measured with the same unit; 3). Fasting glucose- Fasting at least 8 hours prior to measurement of serum glucose (mg/dL). If GLUC was measured in mmol/L, it was requested to convert it into mg/dL, by applying the following coefficient: mmol/l x 18.0182 = mg/dL, based on SBDR - SOCIETY FOR BIOMEDICAL DIABETES RESEARCH; 4). Fasting INS-insulin (µIU/mL). For studies that had measured INS as pmol/L it was asked to convert it to (µIU/mL), by dividing the measure by 6: INS (µIU/mL)=INS (pmol/L) / 6. (See the following paper for why the constant 6 was applied by Heinemann L.[1]); 5). HOMA-B- an indicator of beta-cell function calculated as HOMA-B=(360 x Fasting Insulin) / (Fasting Glucose -60) with fasting insulin expressed in µIU/ml and fasting glucose in mg/dL; 6). HOMA-IR- an indicator of insulin resistance was calculated as HOMAIR=(Fasting Glucose x Fasting Insulin) / 405, with fasting insulin expressed in µIU/mL and fasting glucose in mg/dL; and 7). HbA1c- representing the glycated hemoglobin measured as a percent. For GLUC and INS, cohorts were required to remove individuals that were type 2 diabetics.

All studies transformed variables to the same units, checked the distributions of selected variables and assured they were close to normal with a kurtosis less than 2, and if there were one or two outliers the studies were asked to ensure they were not distorting the distribution of a specific variable. An observation was considered to be an outlier, if it was 1 standard deviation (SD) away from the rest of the points in the trait distribution and beyond 4-6 SDs from the mean. If such a few values existed, then the study was instructed to either winsorize (assign to an outlier a value around 4 SDs) the outlier or set the outlier to a missing value for a specific trait when the study analyst discovered that this outlier value was a data mistake. For BMI, if the data were normal then no changes were to be made, but if the data were non-normal a BLOM transformation was used as the last resort.[2] All studies were instructed to transform INS, HOMA-B, HOMA-IR by natural log. GLUC for each study in its original measure could have been normally distributed. If not, GLUC was not normally distributed, then the studies were instructed to apply a Box-Cox power transformation when deemed appropriate. Only after each study cohort's data were assured to have a close to normal distribution, did the studies performed adjustments. The following covariates were considered for adjusting for any confounding: AGE, AGE^2, SEX (coded male=0, female=1), FC (field center for multi-center studies by creating n-1 dichotomous covariates where n= number of field centers), and PC1, PC2 etc: principal components (PCs) derived using genotyped SNPs: the first PC (and optionally more PCs, if appropriate for African Americans) were instructed to be included. Other additional cohort-specific covariates, if any, were going to be used for controlling additional confounding. The analyses were performed for each specific race separately. The final response variables (Y) were the standardized residuals after any transformation and covariate adjustments. WHR was adjusted in addition for BMI. The studies were requested to exclude individuals younger than 18 years old, without GWAS data, without data for any of the covariates and excluding T2D individuals ONLY from glycemic traits analyses. Participants with missing values for one trait were used for analysis of any other trait for which they had data. The Division of Statistical Genomics at the Washington University in Saint Louis volunteered to serve as a Data Coordinating Center (DCC) by maintaining an anonymous FTP site for studies to perform data

upload, a secured Web page for archiving all meetings, analyses plans, publicly available data-shared including the mitochondria 1000 Genomes reference used for MT imputations, minutes, programs and instructions, and a box.com site shared with specific analysts for downloading single cohorts QC-ed summary results for performing double meta-analyses and QC sharing. Each study cohort reported summary statistics for the phenotypes, which were checked at the DCC for their accuracy (Table S1).

2. **Mitochondrial Single Nucleotide Variant Annotation**

The mitochondrial genome sequence used for annotation of mitochondrial single nucleotide variant and whole genome sequencing studies is controversial[3]. Historically, the first complete sequence of the mitochondrial genome, the Cambridge Reference Sequence, has been used for mitochondrial genome annotation[4]. With improved sequencing, several errors were found in the original sequencing of the Cambridge Reference Sequence[5]. The errors were corrected in the revised Cambridge Reference Sequence (rCRS; GenBank #NC_012920) while maintaining the original nucleotide sequence numbering[5]. The rCRS is from an individual of European origin, mitochondrial haplotype H2a2, which is a much younger mitochondrial genome in the historical time scale. Most genomic studies utilize reference sequences that are the oldest based upon ancestral lineage, which in the case of the mitochondrial genome, is an African haplogroup. Consequently, some microarrays have utilized a Yoruban sequence (AF347015) or a Reconstructed Sapiens Root Sequence (RSRS), a phylogenetically created sequence representing the deepest root [6,7]. However, the Yoruban sequence is 16,571 bp in length and differs from the rCRS at over 40 positions resulting in inconsistencies in the literature and across arrays regarding the nucleotide position of some single nucleotide variants[6]. The RSRS contains three spacers in order to maintain the nucleotide position numbering of the rCRS, but also differs from the rCRS at over 51 sites.

Consequently, all mitochondrial DNA variants from each array (Table S14) were annotated to conform the nucleotide position numbering of the rCRS prior to analyses. The probes used for each microarray were obtained from the manufacturer or dbSNP and aligned to the rCRS using Geneious 8.1[8]. All probes were also submitted through the Standard Nucleotide Basic Local Alignment Search Tool (BLAST) to ensure the probes bound with high specificity (≥90% identity) to the mitochondrial genome. In order to limit any potential binding to nuclear mitochondrial DNA segments, probes that bound to nuclear chromosomes with ≥80% were excluded from all analyses. (Tables S15-S28)

3. **mtDNA Genotyping and Imputation QC**

Each cohort assured that mtDNA variants kept for analyses were of high quality and matching only to mitochondrial DNA. mtDNA markers that were found not valid were dropped from the analyses. The remaining qualitative MT-variants were updated with latest MT Cambridge Revised Sequence (rCRS) of the Human mtDNA positions. The qualitative MT-variants were used as scaffold for haplotype prephasing and for mtDNA imputation. The MT imputation followed the same plan for all cohorts. Each study built genotype data into two files: a pedigree file and a map file known also as PED (pedigree_ID, subject_ID, father_ID, mother_ID, sex, dummy_affection, and Variants (in homozygote letters) in the same order as in the map) and MAP. If any marker was heterozygote, PLINK[9,10] software turned them to missing by default for MT.

We turned files into binary for faster processing (input llfs26.csv; output ganon26):

plink --noweb --file llfs26.csv --out ganon26 --make-bed

We dropped any variants with missengness > 0.9 (input ganon26; output new26):

plink --noweb --bfile ganon26 --geno 0.1 --make-bed --out new26

For prephasing of MT-haplotypes we used SHAPEIT2[11,12] using the following command line (input new26; output ganon26.phased)

shapeit --input-bed ./new26.bed ./new26.bim ./new26.fam --duohmm --rho 4.0E-12 -O ganon26.phased

-duohmm is an option in SHAPEIT program when a cohort contains family data and SHAPEIT will use nuclear family data information for imputation

-rho option was used to define some very tiny recombination rates between variants, because no genetic map was available.

All studies downloaded the following files from 1000 Genomes project, where the first file was to be used as MT-reference in each cohort MT-imputations

ALL.chrMT.phase3_callmom-v0_4.20130502.genotypes.vcf.gz

ALL.chrMT.phase3_callmom-v0_4.20130502.genotypes.vcf.gz.tbi

README_chrMT_phase3_callmom.md

For the needed MT genetic map we created the following file (mtgeneticmap.map):

position COMBINED_rate.cM.Mb. Genetic_Map.cM

1 0 0

16579 0 0.02

The file from 1000G was transformed into OXFORD format file as follows using VCFTOOLS and a new version of PLINK written in C++[10].

vcftools --gzvcf ALL.chrMT.phase3_callmom-v0_4.20130502.genotypes.vcf.gz --max-alleles 2 --min-alleles 2 --max-missing 1 --remove-filtered-geno-all --recode --out chrMTfilter

vcftools --vcf chrMTfilter.recode.vcf --plink-tped --out ALL.chrMT.phase3

plink1_9 --tfile ALL.chrMT.phase3 --recode oxford --out oxg1000ref.gen

sed -i '2d' oxg1000ref.sample

cut -d' ' -f2,3,4,5 < oxg1000ref.gen > oxg1000ref.legend

echo "rsID position a0 a1" > header1.txt

cat header1.txt oxg1000ref.legend > fin.oxg1000ref.legend

The following script was executed to run IMPUTE[13] for imputation based on full MT sequence.

#/bin/bash! impute \ -merge_ref_panels -m ./mtgeneticmap.map -h ./g1000mtref/oxg1000ref.gen \ -l ./g1000mtref/fin.oxg1000ref.legend \ -known_haps_g ../ganon26.phased.haps \ -int 1 16579 \ -Ne 20000 \ -o ./c26_all

The command merge_ref_panels gives the opportunity to use more than one reference, but in our case we used only 1000G MT reference. As expected the IMPUTE software will produce a warning that we are using 1 reference only, as expected.

-int 1 16579 option defined the imputation window, which included the full length of MT-DNA, including all variants for each cohort as well as all markers of 1000G MT reference.

–known_haps_g refers to a cohort specific SHAPEIT prephased haplotypes

The imputed results were in OXFORD format with three probabilities (AA, AB and BB) per cell. Because we wished to use our own programs for MT-variants associations with 7 selected phenotypes, the three probability calls per cell as produced by IMPUTE software had to be transformed into dosage. We used FCGENE software (https://sourceforge.net/projects/fcgene/) for turning three probability calls per cell to dosage for each individual and for each marker imputed:

fcgene --gens ../c26_all \
--thresh 0.9 \
--info ../c26_all_info \
--info-thresh 0.3 \
--maf-thresh 0.00 \
--pedinfo ../ganon26.phased.sample \
--oformat r --transpose \
--force ref-allele=allele2 \
--out ./rtc26

The mtDNA single imputed variants were excluded prior to association tests analyses, when the imputation quality was < 0.30 or they were monomorphic SNVs. The R Package: EasyQC *(V10.0)*[14] (http://www.genepi-regensburg.de/easyqc/) was used at DCC to check for any duplicated results, removing the SNVs with miss-matched alleles against the MT 1000 Genomes reference (phase 3 version 5, 3892 MT-variants, ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502), and eliminating SNVs with allele frequency deviant from the MT 1000 Genomes cosmopolitan reference.

In addition, in house SAS programs (SAS 9.4, SAS Institute Inc., Cary, NC, USA) were used to check the coded alleles and non-coded alleles matching among all cohorts. Any discrepancies of coded alleles were corrected by studies.

The detailed information of the mtDNA genotyping array, the number of mtDNA variants genotyped, the number of mtDNA variants used for imputation, the number of MT variants imputed, imputation accuracy, and the number of mtDNA variants after fcGENE for each cohort are included in Tables S14 and S15-S28.

4. **mtDNA Variants Association Tests (Single variant linear regression and gene based SKAT tests)**

Additional quality control procedures were applied in statistical analysis. These procedures included: 1). removal of mtDNA variants if their imputation quality scores ≤ 0.3; 2). setting mtDNA variants as missing if their dosage values=1; and 3). removal of mtDNA variants if their missing rates ≥ 5%. Heteroplasmic alleles are rare in genotyping of mtDNA variants with genotyping arrays. Therefore, we set these rare heteroplasmic alleles as missing before and after imputation.

For each of the imputed mtDNA variants, the dosage values (0 or 2) were used as the independent variable in association analysis with prepared standardized phenotypes (standardized residuals from normally distributed variables and adjusted for covariates, see 1. Phenotype preparation) as the outcome. An additive genetic model was employed in association analysis using both a self-developed regression model (the linear or linear mixed models written in R programming) and a SKAT (the prepScores() function in seqMeta R package: seqMeta package**:** Meta-Analysis of Region-Based Tests of Rare DNA Variants; https://github.com/DavisBrian/seqMeta) approaches. Familial and maternal correlation structures [15] were accounted for in the analysis of the family data.

For single mtDNA SNV association tests, the models were two:

$\boldsymbol{Y = \beta SNV + \varepsilon}$ (1) for unrelated individuals, where $\boldsymbol{\beta}$ is the estimated slope between SNV dosages regressed on residuals of each of 7 trait responses and $\boldsymbol{\varepsilon}$ is the remaining unexplained effects, and

$\boldsymbol{Y = \beta SNV + \gamma + \delta + \varepsilon}$ (2) for familial, where $\boldsymbol{\gamma}$ is the familial vector of relationships and $\boldsymbol{\delta}$ is an indicator of maternal relationships.

For SKAT gene based mtDNA associations, the models were also two:

$\boldsymbol{Y = GSNVs_i + \varepsilon}$ (3) for unrelated individuals, where $\boldsymbol{G}$ is an n × q genotype matrix for i=1 to q rare SNVs of interest, and

$\boldsymbol{Y = GSNVs_i + \gamma + \delta + \varepsilon}$ (4) for familial designs, where $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are the same as described above in model (2).

5. **mtDNA Meta-analyses (single-variant and gene-based)**

The single mtDNA variant meta-analyses were performed with METAL[16] (Table 1, Figure 1 and Figures S1-S7). For the METAL meta-analysis the Genomic Control correction was set to off, because the MT-DNA it is a small molecule of 16.6KB. The analyses included all ancestries and Europeans only. Filters implemented were: imputation quality >0.3 (implemented at the run of association programs). The threshold used for Bonferroni corrections was considered to be $P \leq$ 1E-3, which is based upon our working group internal permutation tests using the ARIC study MT data, in which 49 MT-DNA variants were considered independent MT-DNA markers. The METAL software used was the 2011-03-25 release.

The possibility that an aggregation of rare or low-frequency alleles in 37 mtDNA coding genes contribute to variation in metabolic phenotypes was tested by the standard burden and sequence

kernel test (SKAT). The standard burden test[17,18] is sensitive in detecting association when all variants have effects on a phenotype in a concordant direction. The SKAT test is designed to detect the effects of alleles that collectively contribute to higher and lower blood pressure. The skatMeta() and burdenMeta() functions in the seqMeta[19] package were used to evaluate aggregate effects at gene-level at minor allele frequency (MAF) ≤0.01 (T1) or MAF ≤0.05 (T5) levels. We used the relaxed threshold $P \leq 0.01$ to demonstrate significance.

6. **mtDNA Genomic Associations**

Previous studies have reported a number of mitochondrial SNVs / genes with pleiotropic effects (a SNV or a gene, or different SNVs within the same gene associate simultaneously with different traits). For example, Hudson *et al*[20] evaluated the association of mtDNA variation in cases/diseases (38,638 individuals with 11 major diseases) and healthy controls (17,483). Among others, they reported association of the MT-3197 (rs2854131) in *MT-RNR2*, the mitochondrial 16S rRNA, with schizophrenia, ulcerative colitis, ankylosing spondylitis, multiple sclerosis, ischemic stroke and Parkinson's disease. The same MT-3197, in our study, was associated with BMI (*P*=2.95E-03, N=127,224, PA), but did not pass our study Bonferroni *P*≤5E-04 threshold. It is possible that mtDNA variants/genes associating with glucose metabolism and insulin signaling may have multiple effects in other disease manifestations beyond T2D. A study by Fetterman *et al*[21] focused on measures of vascular function in Framingham Heart Study - EA in 7,247 individuals genotyped for 268 variants. After multiple corrections, only MT-13966 remained significant in association with PAT ratio (a measure of microvascular function). For the seven traits studied here, no significant associations were found with this variant. Mitchell *et al*[22] focused on cardiovascular related traits and T2D in BioVU - AA 15,863 sample genotyped for 135 variants. The authors of this publication considered their study exploratory in nature and did not correct for multiple testing. They did not report any significant associations with BMI, and the remaining results, including those for T2D, were modest. Another paper was published by Liu *et al*[15]. We summarized the results of three publications mentioned above using a threshold of *P*<5E-02 (Table S13). These comparisons in our findings do not pass our liberal *P*-value (*P*=1E-03) threshold.

7. MT-nDNA Candidate Genes

For identifying MT-nDNA candidate genes we used four sources of candidate genes: a) MitoCarta 2.0[23,24], b) Literature Lab (http://Acumenta.com) , c) MitoMiner (http://mitominer.mrc-mbu.cam.ac.uk/release-4.0/begin.do) and d) MT-defects associated with β-cell dysfunction[25]. MitoCarta 2.0 is an inventory of human and mouse autosomal genes encoding proteins with strong support of mitochondrial localization. To identify mitochondria associated genes, MitoCarta performed mass spectrometry of mitochondria isolated from fourteen tissues. Protein localization was assessed using large-scale GFP tagging/microscopy, and integration of these results with six other genome-scale datasets of mitochondrial localization using a Bayesian approach. The two separate sets (human genes and mouse ortholog genes) from MitoCarta. The second source of autosomal candidate genes was established by us using Literature Lab (Acumenta Biotech, Westminster, MA), a software that mines more than 17.5 million PubMed abstract published since January 1, 1990, using MeSH search terms (MeSH terms are the US National Library of Medicine's controlled vocabulary thesaurus, consisting of sets of terms naming descriptors arranged in both alphabetic and in a hierarchical structure that permits

searching at various levels of specificity). For example, we searched via Literature Lab for MeSH mitochondria and identified 36 terms from which we selected only the 25th upper quartile number of genes from a LPF scoring distribution (Table S29). The term LPF is a quantitative expression of association of number of genes with a term (one of the 36 terms, x/Term1) overlapping (x) with genes in pathway analysis (x/Term2), and expressed as LPF = log (x/T1 * x/T2) (Figure S15). In addition, we kept genes that were from human nomenclature as well as we accepted only genes that had more than 15 abstracts cited per selected gene. The Literature Lab searches identified additional unique genes predicted to have association with mitochondria terms.

We used the software MitoMiner (4.0)[26], which identified additional MT-nDNA candidate genes (filtered with an MT-MitoMiner index ≥0.70, by selecting only terms: "Known mitochondrial" and "Predicted mitochondrial"). In MitoMiner selection, we kept only genes that were from human nomenclature. Finally, we identified a list of genes from a publication for gene expression in mouse, which proteins found in MT were down- or up- regulated in conditions of β-cell dysfunction. Of this list, only 3 were new to MT-candidate genes. The final list of MT-nDNA candidate genes reached 2,283.

We merged the final list of MT-nDNA candidate genes with online available summary meta-results. With the set of nDNA SNPs selected, we queried published GWAS summary results for BMI (Speliotes et al, 2010[27], Yang et al, 2012[28], Berndt et al, 2013[29], Randall et al, 2013[30], Monda et al, 2013[31], Locke et al, 2015[32], Shungin et al, 2015[33], NG et al, 2017[34], Justice et al, 2017[35]); for WHR (Heid et al, 2010[36], Randall et al, 2013[30], Berndt et al, 2013[29], Shungin et al, 2015[33], Justice et al, 2017[35], NG et al, 2017[34], Graph et al, 2017[37]). The summary results data were retrieved from https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files. For glucose metabolism we used the following summary results: GLUC (Saxena et al, 2010[38], Dupuis et al, 2010[39], Manning et al, 2010[40], Scott et al, 2012[41]); for INS (Dupuis et al, 2010[39], Manning et al, 2010[40], Scott et al, 2012[41], Prokopenko et al, 2014[42], Strawbridge *et al*, 2011[43], Walford *et al*, 2016[44]); for HOMAB (Dupuis et al, 2010[39]); for HOMAIR (Dupuis et al, 2010[39]). The summary results data were retrieved from https://www.magicinvestigators.org/downloads/. For HbA1c we used the following resources (Soranzo et al, 2010[45], Wheeler et al, 2017[46]).

Finally, from all these merges, we identified 109 MT-nDNA candidate genes comprising 588 sentinel significant SNPs (one unique per gene and trait combination out of 7 traits) (Table 4).

We have provided detailed analysis in the manuscript for MT-nDNA candidate gene functions. Here in the Supplement, we follow with some additional information. We found many interesting functions within the list of 109 genes: one gene is part of the urea-cycle; one has a role in iron–sulphur cluster biogenesis[47]; four genes regulate intracellular calcium concentration[47]; ten genes are associated with apoptosis[48]; and four genes play a role in production and signaling of reactive oxygen species (ROS)[49] (Figure S11). In our study, two transcription factors *FOXO1* (13q14.11, may play a role in myogenic growth, rs1078892, *P*=5.11E-08) and *FOXO3* (6q21, a possible trigger for apoptosis, rs9400239, *P*=1.61E-08) were associated with BMI, while the deacetylase *SIRT3* (11p15.5, regulates epigenetic gene silencing, rs4758633, *P*=3.44E-10) was associated with HbA1c. They regulate metabolic homeostasis in response to oxidative stress. *SIRT3* deacetylates *FOXO3*[50]. *FOXO1* negatively regulates adipogenesis by binding to the promoter sites of *PPARG* and preventing its transcription, while

insulin represses *FOXO1* action[51,52]. Peserico *et al*[53] found that a low glucose nutrition regimen induced the formation of a complex consisting of *FOXO3*, *SIRT3* and MT-RNA polymerase binding to mtDNA regulatory regions and associated with increased MT-respiration.

8. **Enrichment Analysis of MT-nDNA Candidate Genes**

The 109 MT-nDNA selected genes are candidates for MT based on their protein localization in mitochondrion as well as from mining the published literature. We used Literature Lab and MetaCore software and the corresponding databases to explore the 109 genes. The enrichment analysis provided information of gene relations with terms, pathways, diseases, and gene ontology processes, and clustering to understand MT-nDNA genes' role. Using Literature Lab six gene clusters in association with terms were identified (Figure S10). Clusters, two and five included strong associations with mitochondria and obesity. In Cell-Type-MeSH, the term "Mitochondria" associated with the following genes (% of abstracts in relation with the term and permutation *P*-value): *BCL2* (58), *SIRT3* (8), *OGDH* (6), *PPARG* (6), *HK1* (5), *TOMM40* (4), *TSPO* (4), *CPT1A* (3), *DNM1* (1), *LRPPRC* (1), and *TNF* (1) (*P*=9.8E-03); term "Mitochondria, Liver" associated with *CYP27A1* (18), *CPS1* (16), *OGDH* (15), *CPT1A* (10), *SIRT3* (9), *BCL2* (9), *HADHA* (4), *HK1* (4), *BCKDK* (3), *PPARG* (2) and *LRPPRC*(1) (*P*=9.9E-03); and "Mitochondrial membranes" associated with *TOMM40* (42), *BCL2* (37), *TSPO* (10), *HK1* (3), *PHB2* (2), *DNM1* (2), and *CPT1A* (1) (*P*=1.01E-02). A number of pathways showed significant associations with several of the 109 M-nDAN candidate genes. For example, "Biotransformation" (*P*=0E-04): (The chemical alteration of an exogenous substance by or in a biological system); and terms from Metabolism-MeSH domain such as "Glycolysis" (*P*=5.4E-03), "Citric Acid Cycle" (*P*=7.1E-03), "Carbohydrate metabolism" (*P*=7.7E-03), "Metabolome" (*P*=9.7E-03), and "Lipogenesis" (*P*=1.26E-02) (Table S5).

Using MetaCore, the set of MT-nDNA genes associated (*P*-FDR) with BMI showed enrichment in the pathway of "Protein folding and maturation_POMC processing" (*P*=3.5E-30); process network of "Signal transduction_Neuropeptide signaling pathways" (*P*=1.8E-17); by disease "Musculoskeletal and Neural Physiological Phenomena" (*P*=7.9E-37); "Hyperinsulinism" (*P*=6.8E-35); "Overweight" (*P*=4.2E-34); and "Adrenocortical Hyperfunction" (*P*=1.7E-33). For the set of MT-nDNA candidate genes associated with glycemic traits we found enrichment for disease "Nutritional and Metabolic Diseases" (*P*=2.8E-09) and "Diabetes Mellitus, Type 2" (*P*=1.1E-08). MetaCore produced also significant enrichment results when we submitted the full set of MT-nDNA 109 candidate genes in association with diseases: "Nutrition Disorders" (*P*=5.7E-40); "Depressive Disorder" (*P*=3.9E-37) and "Pituitary Diseases" (*P*=1.3E-36); and in GO Processes with "Chemical homeostasis" (*P*=9.2E-19).

9. **Biological Functional Annotations**

A number of databases such as NCBI gene and dbSNP, UCSC browser, Genecards, 1000 Genomes, GWAS Catalog, Haploreg, RegulomeDB, GTEx, GeneGO, Literature Lab were used for annotating our results. The following software were used for MT-nDNA SNV - gene expression regulation analysis, NCBI Entrez gene (ncbi.nlm.nih.gov/gene/), dbSNP (ncbi.nlm.nih.gov/snp/), HaploReg[54], RegulomeDB[55] and GTEx (gtextportal.org). We performed enrichment analyses via software: GeneGO (portal.genego.com) and Literature Lab (acumenta.com).

The annotation used for the biological inference work were sourced from NCBI dbSNP build 138 (HG19) during the analyses, and updated to dbSNP build 150 (HG38). The importance of the new candidate SNVs / gene lists identified in Tables 1-4 and Tables S6, S8-S9, were mined by means of four methods: enrichment analysis, protein- protein interactions (PPI), analytical gene expression cis-regulation, and analytical gene expression trans-regulation.

The GeneGO/MetaCore (http://thomsonreuters.com/products_services/science/systems-biology/), and Literature Lab of ACUMENTA (http://acumenta.com/) software, (accessed on 11.05. 2017) were used for enrichment analysis. We tested if MT-nDNA genes were significantly enriched among pre-specified gene sets defined in pathways, or by shared roles in particular diseases or biological processes from Gene Ontology. The GeneGO, enrichment analysis consists of matching unique gene symbols of possible targets for the "common", "similar" and "unique" sets with gene symbols in functional ontologies in MetaCore. The probability of a random intersection between a set of gene symbols the size of target list with ontology entities is estimated in p-value of hypergeometric intersection. The lower p-value means higher relevance of the entity to the dataset, which shows in higher rating for the entity.

Literature Lab is a data mining software that searches experimentally-derived gene lists with matches from the scientific literature in a curated vocabulary of 24,000 biological and biochemical terms. It employs statistical and clustering analysis on over 17.5 million PubMed abstracts (from 01/01/1990 to the present) to identify pathways (809 pathways), diseases, compounds, cell biology and other areas of biology and biochemistry. The analysis engine compares statistically the submitted gene set to 1,000 random gene sets generated in the analysis to identify term relationships that are associated with the gene set more than by chance alone.

Furthermore, MT-nDNA candidate SNVs were questioned if they reside in any of regulatory marks, analyzing information from ENCODE and ROADMAP initiatives as summarized by HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php)[56] and RegulomeDB (http://regulome.stanford.edu/)[57].

HaploReg (v.4.1) queries were used to capture functional annotations including the chromatin state segmentation of Ernst et al.[56] and chromatin state segmentation on the Roadmap reference epigenomes, conserved regions by GERP and SiPhy, the narrow peaks called by the ENCODE project on DNAse hypersensitivity experiments, and the SPP narrow peaks called by the ENCODE project on ChIP-seq experiments. HaploReg team has used RefSeq genes from the UCSC Genome Browser and GENCODE for annotation. BEDTools were used to calculate the proximity of each variant to a gene by either annotation, as well as the orientation (3' or 5') relative to the nearest end of the gene, based on the strand of the gene. This software was accessed on 11.05.2017 (Table S9).

RegulomeDB was used to summarize the following data types, transcription factor binding sites, ChIP factors: 740 unique data sets including most recent ENCODE data release[55,58]. RegulomeDB uses the Position-Weight Matrix for TF binding, and databases JASPAR CORE, TRANSFAC and UniPROBE[59]. For the DNase sensitivity it uses 204 unique datasets including most recent ENCODE data release (ENCODE Project Consortium). For the Chromatin States, RegulomeDB extracts information from Roadmap Epigenome Consortium from 127 standard epigenomes. Further, RegulomeDB reports eQTLs from several tissue types and also reports DNase footprinting[57,60], differentially methylated regions[61], manually curated regions and validated functional SNVs. RegulomeDB produces a ranking score in which smaller values are

better. RegulomeDB scoring has classes defined as 1b, 1d and 1f: likely to affect binding and linked to expression of a gene target as: 1b (eQTL + TF binding + any motif + DNase footprint + DNase peak); 1d (eQTL + TF binding + any motif + DNase peak); 1f (eQTL + TF binding/DNase peak), 2a and 2b: likely to affect binding, 3a: less likely to affect binding, 4, 5, and 6: minimal binding evidence, and 7: no data. This software was accessed on 11.05.2017 (Table S9).

We also gathered evidence for eQTLs based on GTExportal.org, GRASP software and special gene expression reported results (Westra [62], Lappalainen 2013 [63]) (Table S10). We used also GTEx and Protein Atlas (proteinatlas.org) for gene expression profiles of mtDNA genes (Figure S13).

GWAS3D (http://jjwanglab.org/gwas3d)[64] an online software was used to analyze genetic variants that could affect regulatory elements, by integrating annotations from cell type-specific chromatin states, epigenetic modifications, sequence motifs and cross-species conservation. The regulatory elements are inferred from the genome-wide chromosome interaction data, chromatin marks in 16 different cell types measured by high-throughput chromosome conformation capture technologies (5C, ChIA-PET and Hi-C) from the ENCODE project, Gene Expression Omnibus (GEO) database, published resources and 73 regulatory factor motifs from the Encyclopedia of DNA Element project (Figures S12.1-S12.3). This software was accessed on 01.10.2017.

The importance of novel MT candidate genes were evaluated via protein-protein interactions (PPI) of databases BIND (http://bind.ca), BioGrid (http://thebiogrid.org/), EcoCyc (http://www.ecocyc.org) and HPRD (http://www.hprd.org/) as summarized by NCBI ftp://ftp.ncbi.nih.gov/gene/GeneRIF/) (accessed on 02.11. 2018). The PPI is a topological summary of all known or predicted protein interactions, based on stable physical associations, transient binding, substrate chaining, information relay, sourced from curated experimental data of biochemical, biophysical and genetic techniques, and predicted computationally. The importance of candidate gene list based on PPI was evaluated using igraph package (http://igraph.org)[65]. The network was built using our programs in SAS, to a Pajek format and imported into igraph in R language. “Google” PageRank algorithm that evaluates the importance of pages/genes in a network was also implemented by igraph[66]. The graph figure of this network is not reported, because it is a very large network.

Because in our study sample sizes per mtDNA marker, were variable and quite often small, we did not address the association of significant markers by gender, which future studies such as TopMed initiative of NHLBI with full sequencing of nDNA and mtDNA may deem it appropriate to contribute in the future.

4. References

1. Heinemann, L. Insulin assay standardization: leading to measures of insulin sensitivity and secretion for practical clinical care: response to Staten et al. *Diabetes Care* **33**, e83; author reply e84 (2010).
2. Beasley, T.M., Erickson, S. & Allison, D.B. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav Genet* **39**, 580-95 (2009).
3. Bandelt, H.J., Kloss-Brandstatter, A., Richards, M.B., Yao, Y.G. & Logan, I. The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies. *J Hum Genet* **59**, 66-77 (2014).
4. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-65 (1981).
5. Andrews, R.M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147 (1999).
6. Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-13 (2000).
7. Behar, D.M. *et al.* A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet* **90**, 675-84 (2012).
8. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-9 (2012).
9. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
10. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
11. O'Connell, J. *et al.* A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet* **10**, e1004234 (2014).
12. Delaneau, O., Marchini, J., Genomes Project, C. & Genomes Project, C. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun* **5**, 3934 (2014).
13. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457-70 (2011).
14. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-212 (2014).
15. Liu, C., Dupuis, J., Larson, M.G. & Levy, D. Association testing of the mitochondrial genome using pedigree data. *Genet Epidemiol* **37**, 239-47 (2013).
16. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
17. Morgenthaler, S. & Thilly, W.G. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res* **615**, 28-56 (2007).
18. Li, B. & Leal, S.M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* **83**, 311-21 (2008).
19. Wu, M.C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82-93 (2011).
20. Hudson, G., Gomez-Duran, A., Wilson, I.J. & Chinnery, P.F. Recent mitochondrial DNA mutations increase the risk of developing common late-onset human diseases. *PLoS Genet* **10**, e1004369 (2014).
21. Fetterman, J.L. *et al.* Relations of mitochondrial genetic variants to measures of vascular function. *Mitochondrion* (2017).

22. Mitchell, S.L. *et al.* Investigating the relationship between mitochondrial genetic variation and cardiovascular-related traits to develop a framework for mitochondrial phenome-wide association studies. *BioData Min* **7**, 6 (2014).
23. Calvo, S.E., Clauser, K.R. & Mootha, V.K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res* **44**, D1251-7 (2016).
24. Pagliarini, D.J. *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112-23 (2008).
25. Lu, H., Koshkin, V., Allister, E.M., Gyulkhandanyan, A.V. & Wheeler, M.B. Molecular and metabolic evidence for mitochondrial defects associated with beta-cell dysfunction in a mouse model of type 2 diabetes. *Diabetes* **59**, 448-59 (2010).
26. Smith, A.C. & Robinson, A.J. MitoMiner v3.1, an update on the mitochondrial proteomics database. *Nucleic Acids Res* **44**, D1258-61 (2016).
27. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-48 (2010).
28. Yang, J. *et al.* FTO genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267-72 (2012).
29. Berndt, S.I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* **45**, 501-12 (2013).
30. Randall, J.C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* **9**, e1003500 (2013).
31. Monda, K.L. *et al.* A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nat Genet* **45**, 690-6 (2013).
32. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
33. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187-196 (2015).
34. Ng, M.C.Y. *et al.* Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. *PLoS Genet* **13**, e1006719 (2017).
35. Justice, A.E. *et al.* Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun* **8**, 14977 (2017).
36. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* **42**, 949-60 (2010).
37. Graff, M. *et al.* Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. *PLoS Genet* **13**, e1006528 (2017).
38. Saxena, R. *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* **42**, 142-8 (2010).
39. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-16 (2010).
40. Manning, A.K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* **44**, 659-69 (2012).
41. Scott, R.A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* **44**, 991-1005 (2012).
42. Prokopenko, I. *et al.* A central role for GRB10 in regulation of islet function in man. *PLoS Genet* **10**, e1004235 (2014).

43. Strawbridge, R.J. *et al.* Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624-34 (2011).
44. Walford, G.A. *et al.* Genome-Wide Association Study of the Modified Stumvoll Insulin Sensitivity Index Identifies BCL2 and FAM19A2 as Novel Insulin Sensitivity Loci. *Diabetes* **65**, 3200-11 (2016).
45. Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229-39 (2010).
46. Wheeler, E. *et al.* Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med* **14**, e1002383 (2017).
47. Tuppen, H.A., Blakely, E.L., Turnbull, D.M. & Taylor, R.W. Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta* **1797**, 113-28 (2010).
48. Newmeyer, D.D. & Ferguson-Miller, S. Mitochondria: releasing power for life and unleashing the machineries of death. *Cell* **112**, 481-90 (2003).
49. Starkov, A.A. The role of mitochondria in reactive oxygen species metabolism and signaling. *Ann N Y Acad Sci* **1147**, 37-52 (2008).
50. Tseng, A.H., Wu, L.H., Shieh, S.S. & Wang, D.L. SIRT3 interactions with FOXO3 acetylation, phosphorylation and ubiquitinylation mediate endothelial cell responses to hypoxia. *Biochem J* **464**, 157-68 (2014).
51. Farmer, S.R. The forkhead transcription factor Foxo1: a possible link between obesity and insulin resistance. *Mol Cell* **11**, 6-8 (2003).
52. Armoni, M. *et al.* FOXO1 represses peroxisome proliferator-activated receptor-gamma1 and -gamma2 gene promoters in primary adipocytes. A novel paradigm to increase insulin sensitivity. *J Biol Chem* **281**, 19881-91 (2006).
53. Peserico, A. *et al.* A novel AMPK-dependent FoxO3A-SIRT3 intramitochondrial complex sensing glucose levels. *Cell Mol Life Sci* **70**, 2015-29 (2013).
54. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**, D930-4 (2012).
55. Xie, D. *et al.* Dynamic trans-acting factor colocalization in human cells. *Cell* **155**, 713-24 (2013).
56. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-9 (2011).
57. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-7 (2012).
58. Boyle, A.P. *et al.* Comparative analysis of regulatory information and circuits across distant species. *Nature* **512**, 453-6 (2014).
59. Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152**, 327-39 (2013).
60. Degner, J.F. *et al.* DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390-4 (2012).
61. Kuleshov, V. *et al.* Whole-genome haplotyping using long reads and statistical methods. *Nat Biotechnol* **32**, 261-6 (2014).
62. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-43 (2013).
63. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506-11 (2013).

64. Li, M.J., Wang, L.Y., Xia, Z., Sham, P.C. & Wang, J. GWAS3D: Detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. *Nucleic Acids Res* **41**, W150-8 (2013).
65. Csardi, G. & Nepusz, T. The igraph software package for complex network research, InterJournal, Complex Systems 1695. (2006).
66. Networks, Crowds, and Markets. Reasoning about a Highly Connected World. David Easley and Jon Kleinberg. Cambridge University Press. 351-383 (2010).

5. CHARGEmtDNA+ Study Acknowledgments

Study	Acknowledgment
AIRWAVE	We thank all participants in the Airwave Health Monitoring Study. The Study is funded by the Home Office (grant number 780-TETRA) with additional support from the National Institute for Health Research (NIHR), Imperial College Healthcare NHS Trust (ICHNT) and Imperial College Biomedical Research Centre (BRC). P.E. acknowledges support from the ICHNT and Imperial College BRC, the MRC-PHE Centre for Environment and Health (MR/L01341X/1), and the NIHR Health Protection Research Unit on Health Impact of Environmental Hazards (HPRU-2012-10141). This work used the computing resources of the UK MEDical BIOinformatics partnership programme (UK MED-BIO) (MR/L01632X/1). This work was supported by the UK Dementia Research Institute which receives its funding from UK DRI Ltd funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK.
ALSPAC (The Avon Longitudinal Study of Parents and Children)	We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.
ARIC (The Atherosclerosis Risk in Communities Study)	The ARIC study is a population-based cohort study designed to study new and established risk factors for atherosclerosis and community trends in coronary heart disease. In 1987-89, baseline data was collected on 15,792 adults, aged 45–64 y, living in four U.S. communities (Forsyth County, NC; Jackson, MS; northwest Minneapolis suburbs, MN; Washington County, MD). The baseline exam was conducted in 1987-89 and information was collected on African Americans, Caucasians, and a few adults of other ethnicities, aged 45–64 y. After providing informed consent, 15,792 adults were enrolled (8,710 women and 7,082 men). Up to 8,591, Caucasian adults with available DNA, valid dietary information, and consent to share genetic data were eligible for the current analysis. The Atherosclerosis Risk In Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Dr. Nettleton is supported by a K01 from the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases (5K01DK082729-04).
Danish: The 1936 Birth Cohort / 60 Years	The collection of the cohort was financially supported by The Danish Heart Foundation and The Danish Medical Research Council. The authors are grateful to the staff at the Centre of Preventive Medicine, and to MD, general practitioner, Professor Hanne Hollnagel Dr Med. Sci., who initiated the study of the 1936 cohort.

Danish: Addition-Pro	The ADDITION-PRO study was funded by an unrestricted grant from the European Foundation for the Study of Diabetes/Pfizer for Research into Cardiovascular Disease Risk Reduction in Patients with Diabetes (74550801), the Danish Council for Strategic Research, internal research and equipment funds from Steno Diabetes Center and supported by research grants from the Novo Nordisk Foundation. The ADDITION-PRO study is managed by the ADDITION-DK steering committee (Torsten Lauritzen, Knut Borch-Johnsen, Annelli Sandbæk, Marit E. Jørgensen, and Daniel Witte).
Danish: Cogen	The Copenhagen Cardiovascular Genetic study (COGEN) is a biobank that has collected superfluous whole blood from patients admitted to six cardiology departments in the greater region of Copenhagen from 2010-2017. COGEN currently contains samples from ~80,000 individuals. Due to a Danish permanent identification number (given to all permanent residents), these data can be linked with various clinical databases at an individual level. The present study population is comprised of individuals who have had at least one coronary angiogram performed between 2010-2014. Data on the angiograms were collected from the Eastern Danish Heart Registry, which is a clinical database, where information on demographics (e.g. age, sex), risk factors and comorbidities (e.g., smoking status, diabetes) has been routinely entered on all patients in the Eastern region of Denmark. All data were de-identified prior to analyses. The ethics committee of Region North Jutland (N-20140048) approved the project and COGEN has permission from the Data Protection Agency (00916 GEH-2010-001).
Danish: Goya	The GOYA study was conducted as part of the activities of the Danish Obesity Research Centre (DanORC, www.danorc.dk) and the MRC centre for Causal Analyses in Translational Epidemiology (MRC CAiTE), and genotyping was funded by the Wellcome Trust (WT 084762MA). GOYA is a nested study within The Danish National Birth Cohort which was established with major funding from the Danish National Research Foundation. Additional support for this cohort has been obtained from the Pharmacy Foundation, the Egmont Foundation, The March of Dimes Birth Defects Foundation, the Augustinus Foundation, and the Health Foundation.
Danish: Health2006	The Health2006 study was financially supported by grants from the Velux Foundation; the Danish Medical Research Council, Danish Agency for Science, Technology and Innovation; the Aase and Ejner Danielsens Foundation; ALK-Abelló A/S (Hørsholm, Denmark), Timber Merchant Vilhelm Bangs Foundation, MEKOS Laboratories (Denmark) and Research Centre for Prevention and Health, the Capital Region of Denmark.
Danish: Inter99	The Inter99 study is funded by The Danish Medical Research Council, The Danish Centre for Evaluation and Health Technology Assessment, Novo Nordisk, Copenhagen County, The Danish Heart Foundation, The Danish Pharmaceutical Association, Augustinus foundation, Ib Henriksen foundation and Becket foundation. The staff from Research Centre for Prevention and Health, The capital region, Glostrup, Denmark is acknowledged their effort in making the Inter99 study possible. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (www.metabol.ku.dk).
Danish: Vejle Biobank	The project was funded by the Danish Council for Independent Research/Medical Sciences, the Research Council of Vejle Hospital, the Department of Internal Medicine, Vejle Hospital, Vejle County, the Danish Research Fund, the Lions Club International Denmark, and anonymous donations. The authors would like to thank the laboratory technologists engagement and dedicated work.
EPIC-InterAct	We thank all EPIC participants and staff for their contribution to the EPIC-InterAct study. Funding for the EPICInterAct project was provided by the EU FP6 programme (LSHM_CT_2006_037197).

EPIC-Norfolk	EPIC-Norfolk is supported by programme grants from the Medical Research Council UK (G1000143) and Cancer Research UK (C864/A14136) and with additional support from the European Union, Stroke Association, British Heart Foundation, Research into Ageing, Department of Health, The Wellcome Trust and the Food Standards Agency. NJW and CL also acknowledge support from the Medical Research Council, UK (MC_UU_12015/1; MC_PC_13048). We thank all EPIC participants and staff for their contribution to the study, and the laboratory teams at the MRC Epidemiology Unit for sample management.
FAMHS (Family Heart Study)	This research was conducted using data and resources form the NHLBI Family Heart Study and Washington University School of Medicine. This work was partially supported by the NIDDK R01DK089256 and NHLBI R01HL117078. We would also like to thank the families that participated in the Family Heart Study.
Fenland Study	The Fenland Study is funded by the Wellcome Trust and the Medical Research Council (MC_U106179471). We are grateful to all the volunteers for their time and help, and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams. We further acknowledge support from the Medical research council (MC_UU_12015/1).
FHS (Framingham Heart Study)	The Framingham Heart Study is funded by National Institutes of Health contract N01-HC-25195. This work is also supported by National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) U01 DK078616 and NIDDK K24 DK080140. In addition, analysis of FHS data and meta-analysis of the consortium's data supported by NHLBI intramural research funding. The NHLBI grant R01 HL105756 contributed in the CHARGE infrastructure making possible the Working Group collaborations.
GeneSTAR (Genetic Study of Atherosclerosis Risk)	GeneSTAR was supported by grants from the National Institutes of Health/National Heart, Lung, and Blood Institute (U01 HL72518, HL087698, HL49762, HL59684, HL58625, HL071025, and HL092165), National Institutes of Health/National Institute of Nursing Research (NR0224103 and NR008153), National Institutes of Health/National Institute of Neurological Disorders and Stroke (NS062059), and by a grant from the National Institutes of Health/National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center.
GENOA (Genetic Epidemiology Network of Arteriopathy)	Funding: This work was supported by United States Public Health Service Grants from the National Institutes of Health (K23 DK78229, R01 DK73537 and R01 HL 87660).
GOLDN (Genetics of Lipid Lowering Drugs and Diet Network Study)	GOLDN authors wish to thank all of the families who contributed to the study and all investigators who participated in this effort. GOLDN participants were recruited and phenotyped under NIH National Heart, Lung and Blood Institute (NHLBI) grant U01HL072524. Genome-wide common variant data were collected under NHLBI grant R01091357. Genome-wide methylation data were collected under NHLBI grant R01104135.

HCHS-SOL (The Hispanic Community Health Study/Study of Latinos)	The Hispanic Community Health Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago – HHSN268201300003I / N01-HC-65236 Northwestern University), and San Diego State University (HHSN268201300005I / N01-HC-65237). The following Institutes/Centers/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03). Additional analysis support was provided by NIDDK grant 1R01DK101855-01, NHLBI grant N01HC65233, and AHA grant 13GRNT16490017. Genotyping efforts were supported by the NIH Department of Health and Human Services (HSN26220/20054C), National Center for Advancing Translational Science Clinical Translational Science Institute (UL1TR001881), and NIDDK Diabetes Research Center (DK063491). This manuscript has been reviewed by the HCHS/SOL Publications Committee for scientific content and consistency of data interpretation with previous HCHS/SOL publications.
HYPERGEN (Hypertension Genetic Epidemiology Network)	We thank all individuals who participated in the Hypertension Genetic Epidemiology Network study. The HyperGEN study is funded by cooperative agreements (U10) with NHLBI: HL54471, HL54472, HL54473, HL54495, HL54496, HL54497, HL54509, HL54515, and 2 R01 HL55673-12. The study involves: University of Utah: (Network Coordinating Center, Field Center, and Molecular Genetics Lab); Univ. of Alabama at Birmingham: (Field Center and Echo Coordinating and Analysis Center); Medical College of Wisconsin: (Echo Genotyping Lab); Boston University: (Field Center); University of Minnesota: (Field Center and Biochemistry Lab); University of North Carolina: (Field Center); Washington University: (Data Coordinating Center); Weil Cornell Medical College: (Echo Reading Center); National Heart, Lung, & Blood Institute. For a complete list of HyperGEN Investigators: http://www.biostat.wustl.edu/hypergen/Acknowledge.html
LBC1921 and LBC1936 (Lothian Birth Cohort 1921 Lothian Birth Cohort 1936)	Phenotype collection in the Lothian Birth Cohort 1921 was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Genotyping was supported by Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, and the Royal Society of Edinburgh. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and Medical Research Council (MRC) is gratefully acknowledged. WDH is supported by a grant from Age UK (Disconnected Mind Project).
LLFS (Long Life Family Study)	We would like to thank the staff and participants of the Long Life Family Study for making this work possible. LLFS was supported by the National Institute on Aging (U01-AG023712, U01-AG23744, U01-AG023746, U01-AG023749, and U01-AG023755).
MESA (Multi-Ethnic Study of Atherosclerosis)	This research, including GWAS and Cardiometabochip genotyping, was supported part by the Multi-Ethnic Study of Atherosclerosis (MESA) contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. Funding for MESA Family was provided by grants R01HL071051, R01HL98077, R01HL071205, R01HL071250, R01HL071251, R01HL071258, R01HL071259, P50 ES015915, and by the

	National Center for Research Resources, and Grant UL1RR033176. Funding for MESA Share genotyping was provided by NHLBI Contract N02-HL-6-4278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research (CRC) grant DK063491. Although the research described in this presentation has been funded in part by the United States Environmental Protection Agency through RD831697 to the University of Washington, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.
NEO (The Netherlands Epidemiology of Obesity study)	The authors of the NEO study thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Pat van Beelen, Petra Noordijk and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. Diana van Heemst is supported by the European Commission funded project HUMAN (Health-2013-INNOVATION-1-602757). Dennis Mook-Kanamori is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023).
OOA (The Old Order Amish Study)	The authors thank Dr. Braxton D. Mitchell and Dr. Alan R. Shuldiner for their contribution in Amish study and the Amish Research Clinic Staff for their great efforts in study subject recruitment and characterization as well as all the volunteers in the Amish community for their participation in these studies. The Amish portion of this study was supported by NIH grants R01 HL121007, U01 HL072515, R01 AG18728, P30 DK072488, and AHA17GRNT33440151.
PELOTAS (1982 Pelotas (Brazil) Birth Cohort Study)	The 1982 Pelotas Birth Cohort Study is conducted by the Postgraduate Program in Epidemiology at Universidade Federal de Pelotas with the collaboration of the Brazilian Public Health Association (ABRASCO). From 2004 to 2013, the Wellcome Trust supported the study. The International Development Research Center, World Health Organization, Overseas Development Administration, European Union, National Support Program for Centers of Excellence (PRONEX), the Brazilian National Research Council (CNPq), and the Brazilian Ministry of Health supported previous phases of the study. Genotyping of 1982 Pelotas Birth Cohort Study participants was supported by the Department of Science and Technology (DECIT, Ministry of Health) and National Fund for Scientific and Technological Development (FNDCT, Ministry of Science and Technology), Funding of Studies and Projects (FINEP, Ministry of Science and Technology, Brazil), Coordination of Improvement of Higher Education Personnel (CAPES, Ministry of Education, Brazil).
RS (ROTTERDAM STUDY II & III)	The RS is supported by the Erasmus MC University Medical Center and Erasmus University Rotterdam; The Netherlands Organization for Scientific Research (NWO); The Netherlands Organization for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); The Netherlands Genomics Initiative (NGI); the Ministry of Education, Culture and Science; the Ministry of Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam. The contribution of inhabitants, general practitioners and pharmacists of the Ommoord district to the Rotterdam Study is gratefully acknowledged.

SHIP (Study of Health In Pomerania)	We thank all staff members as well as the genotyping staff involved in the generation of the SNP data. SHIP (Study of Health in Pomerania) and SHIP-TREND both represent population-based studies. SHIP is supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung (BMBF); grants 01ZZ9603, 01ZZ0103, and 01ZZ0403) and the German Research Foundation (Deutsche Forschungsgemeinschaft (DFG); grant GR 1912/5-1). SHIP and SHIP-TREND are part of the Community Medicine Research net (CMR) of the Ernst-Moritz-Arndt University Greifswald (EMAU) which is funded by the BMBF as well as the Ministry for Education, Science and Culture and the Ministry of Labor, Equal Opportunities, and Social Affairs of the Federal State of Mecklenburg-West Pomerania. The CMR encompasses several research projects that share data from SHIP. The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH. SNP typing of SHIP and SHIP-TREND using the Illumina Infinium HumanExome BeadChip (version v1.0) was supported by the BMBF (grant 03Z1CN22).
TUDR (Taiwan-US Diabetic Retinopathy Study)	This study was supported by the National Eye Institute of the National Institutes of Health (EY014684 to J.I.R. and Y.-D.I.C.) and ARRA Supplement (EY014684-03S1, -04S1), the National Institute of Diabetes and Digestive and Kidney Disease grant DK063491 to the Southern California Diabetes Endocrinology Research Center, the Eye Birth Defects Foundation Inc., the National Science Council, Taiwan (NSC 98-2314-B-075A-002-MY3 to W.H.S.) and the Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1003001C to W.H.S .). DNA handling and genotyping were supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881 and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
WGHS (Women's Genome Health Study)	The WGHS is supported by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute (CA047988 and UM1CA182913), with collaborative scientific support and funding for genotyping provided by Amgen.
WHI (The Women's Health Initiative)	Funding support for the "Epidemiology of putative genetic variants: The Women's Health Initiative" study is provided through the NHGRI grants HG006292 and HL129132. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSC271201100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf.
WUHHD (Washington University Hypertensive Heart Disease)	We thank all individuals who participated in the Washington University Hypertensive Heart Disease Study and to Alan D. Waggoner for his contributions to this study. The WU-HHD study was supported by National Institutes of Health (NIH) Grants R01HL091028, R01HL071782, T32HL083822, R21HL094668, UL1RR024992 and KL2RR024994 (Washington University), and an American Heart Association grant #0855626G. Additional support provided by NIH grant S10RR024532, a grant from the Barnes-Jewish Hospital Foundation, and a Lucille P. Markey Fellowship.

6. CHARGEmtDNA+ Working Group Study Descriptions

AIRWAVE (The Airwave Study)

The Airwave Health Monitoring Study is a cohort of UK police officers and staff across Great Britain. The study design and rationale have previously been described[1]. Briefly, participants from each force who agreed to participate were enrolled from 2004 either with an enrolment questionnaire or a comprehensive health screening performed locally. At the health screen, the participant also filled out an extensive questionnaire on a touchscreen computer. Both questionnaires include demographic, health and lifestyle questions, and information on TETRA radio usage. The time between the enrolment questionnaire and the health screening was determined by logistic constraints and varied between 6 months and one or more years. As of 31 December 2012, the Airwave Health Monitoring Study had enrolled 42,112 participants. For participating forces, on average, 50% of their employees were recruited once enrolment was complete. Participants signed a consent form permitting use of their data and samples for future research. The study had ethical approval through the National Health Service multi-site research ethics committee (MREC/13/NW/0588).

The individuals were genotyped using Illumina Infinium HumanCoreExome-12 v1.1 BeadChip Array. We extracted 278 mitochondrial markers and updated with latest Cambridge Revised Sequence (rCRS) of the Human MT-DNA positions and were used for haplotype prephasing and imputation. PLINK was used to prepare the data in binary format and 20 variants were excluded due to missing call rates > 0.01. We used SHAPEIT2 for prephasing haplotype scaffold. Imputation was done in one window in the length of the full mitochondrial genome using IMPUTE2 based on the Cosmopolitan 1000 Genomes Phase 3 mitochondrial DNA of 2505 individuals from 26 populations conforming with the rCRS of the Human MT-DNA positions known also as v.5 phase 3 of 1000 Human Genomes HG19, GRCh37 cosmopolitan mitochondrial reference panel (ftp.1000genomes.ebi.ac.uk/vol1/ftp/) released in 2016. The final imputed dosage file (after fcGENE converting and filtering) contained 3399 mitochondrial variants for 14,688 subjects. Based on the IMPUTE2 imputation accuracy assessment, the imputation concordance rate between original mitochondrial genotypes and the same imputed ones for Airwave was %98.9. A few heterozygotes in the original data were considered as possible genotyping errors and turned to missing via PLINK.

ALSPAC (The Avon Longitudinal Study of Parents and Children)

ALSPAC is a population-based cohort study investigating genetic and environmental factors that affect the health and development of children. The study methods are described in detail elsewhere [2,3] (http://www.bristol.ac.uk/alspac). Briefly, 14,541 pregnant women residents in the former region of Avon, UK, with an expected delivery date between 1st April 1991 and 31st December 1992, were eligible to take part in ALSPAC. Detailed information and biosamples have been collected on these women and their offspring at regular intervals, which are available through a searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/).

Written informed consent was obtained for all study participants. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

Genotyping and Imputation: Genotype data was generated using the Illumina HumanHap550 quad genome-wide SNP genotyping platform (Illumina Inc, San Diego, USA) by the Wellcome Trust Sanger Institute (WTSI, Cambridge, UK) and the Laboratory Corporation of America (LCA, Burlington, NC, USA). Samples were excluded based on incorrect sex assignment; abnormal heterozygosity (<0.320 or >0.345 for WTSI data; <0.310 or >0.330 for LCA data); high missingness (>3%); cryptic relatedness (>10% identity by descent) and non-European ancestry (detected by multidimensional scaling analysis).

After QC, 500,527 SNP loci were available for the directly genotype dataset on 8,365 unrelated individuals. Imputation was undertaken using MACH 1.0.16 Markov Chain haplotyping software [4], using CEPH individuals from phase 2 of the HapMap project (hg18) as a reference set (release 22). All variants were filtered to have Hardy-Weinberg equilibrium $P > 5\times10^{-7}$ and an imputation quality score ≥ 0.8 or higher. The positions of the remaining 33 MT markers were updated using the latest MT Cambridge Revised Sequence (rCRS) for Human MT-DNA in preparation for the CHARGEmtDNA+ working group imputation protocol.

Phenotypes: Phenotype data were obtained from various time points in the ALSPAC cohort as some traits were only measured at certain ages in the offspring. At the age 17 clinic (mean age: 17.8, standard deviation: 0.4), height was measured to the nearest 0.1cm using a Harpenden stadiometer (Holtain Crosswell, Dyfed, UK) and weight was measured to the nearest 0.1kg using Tanita electronic scales. Body Mass Index (BMI) was calculated as $(\text{weight (kg)})/(\text{height (m)})^2$. At the age 11 clinic (mean age: 11.7, standard deviation: 0.2), waist and hip circumference were measured to the nearest 0.1 centemeter (cm) and used to calculate waist-hip ratio (WHR).

At the age 15 clinic (mean age: 15.5, standard deviation: 0.3), blood samples were obtained from participants who were either asked to fast overnight (for those attending in the morning) or for a minimum of 6 hours for those attending after lunch. Blood samples were immediately spun and frozen at –80^{o}C and assayed in batches shortly after (3-12 months). Insulin was measured by automated microparticle enzyme immunoassay that does not cross-react with proinsulin and plasma glucose was measured by automated enzymatic. Glucose measurements were obtained by the staff of the Routine Lipids Section of the Biochemistry Department of Glasgow Royal Infirmary using a Hitachi Modular p analyser, (with enzymatic colorimetic assay) and kit supplied by Roche Diagnostics. HbA_{1C} was measured at approximately age 10 (mean age: 10.1, standard deviation: 0.4) from non-fasting venous whole blood and stored in HPLC haemolysant at -70^{o}C for up to 1 year prior to assay. Measurements were obtained using an ion-exchange HPLC assay using the HA-8140 Hi-Auto HbA_{1c} analyser (Menarini Diagnostics), maintained in alignment with the Diabetes Control and Complications Trial (DCCT) method.

ARIC (The Atherosclerosis Risk in Communities Study)

The ARIC study [5], sponsored by the National Heart, Lung and Blood Institute (NHLBI), is a prospective epidemiologic study conducted in four U.S. communities: Minneapolis, MN, Washington County, MD, Forsyth County, NC, and Jackson, MS. ARIC is designed to investigate the etiology and natural history of atherosclerosis, the etiology of clinical atherosclerotic diseases, and variation in cardiovascular risk factors, medical care and disease by

race, gender, location, and date. The ARIC Cohort Component began in 1987, and each ARIC field center randomly selected and recruited a cohort sample of approximately 4,000 individuals aged 45-64 from a defined population in their community. A total of 15,792 participants (~27% African American; ~55% women) received an extensive examination, including medical, social, and demographic data. These participants were reexamined every three years with the first screen (baseline) occurring in 1987-89, the second in 1990-92, the third in 1993-95, and the fourth in 1996-98, and the fifth in 2011-2013. ARIC is an ongoing study, and contemporary interests include aspects of heart failure and healthy aging. Phenotypes for the current study were taken from the baseline visit (visit 1) during 1987-1989. After exclusions for those missing phenotype and genotype data, a maximum of 9481 European and 2860 African ancestry individuals were included in analyses.

Genotyping process

DNA samples were genotyped using the Affymetrix 6.0 SNP array and the Birdseed calling algorithm at the Broad Institute Center for Genotyping and Analysis (CGA). Genotypic data that passed initial quality control at the Broad Center for Genotyping and Analysis (CGA) were released to the GENEVA Coordinating Center (CC), the NCBI dbGaP team and the ARIC project team. For European ancestry samples, a total of 9743 individuals were genotyped. Of these 402 individuals were dropped for first degree relatedness, unexpected duplicates, PC outliers, and sex discrepancies. A total of 9345 European ancestry samples were further used for imputation and other analyses. For African ancestry samples, a total of 3207 individuals were genotyped. Of these 333 individuals were dropped for first degree relatedness, lack of consistency with prior genotypes, and PC outliers. A total of 2874 African ancestry samples were further used for imputation and other analyses. A total of sixty variants on the mitochondria were available for imputation for European and African ancestry individuals. The positions of these markers was updated using the latest MT Cambridge Revised Sequence (rCRS) for Human MT-DNA in preparation for the CHARGEmtDNA+ working group imputation protocol. We used PLINK software to prepare data in binary format, SHAPEIT2 for pre-phasing mtDNA haplotypes and IMPUTE2 combined with the latest 1000 Human Genomes Reference (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/), April 2016 release, for performing imputation of the full mitochondrial genome. In AA, the imputation quality was 99.3%, yielding 3622 mtDNA variants. In EA, imputation quality was 99.4%, yielding 3622 mtDNA variants. After applying fcGENE to convert probabilities per cell to dosage, 450 mtDNA variants in AA and 205 mtDNA variants in EA remained polymorphic.

The 1936 Birth Cohort / 60 Years

The cohort consists of 656 mostly non-diabetic individuals born in 1936, who, on 2 April 1976, were resident in one of four municipalities nearby Glostrup Hospital, Denmark. The cohort was collected to assess the age-specific prevalence of diabetes mellitus and impaired glucose tolerance in 60-year-old individuals in 1996/97 [6].

Addition-Pro

ADDITION-PRO is a longitudinal cohort study of individuals at low to high risk of type 2 diabetes, nested within the population-based ADDITION-Denmark trial of screen-detected diabetes. The ADDITION-PRO cohort comprises 2082 adults (>45 years) collected to have IGT, IFG, or NGT either with high or low risk of developing type 2 diabetes (based on

information about age, sex, gestational diabetes, family history of diabetes, hypertension, BMI, and level of physical activity). 1548 genotyped individuals was included in the present study. The samples were collected in 2009–2011from four Danish research centers (Steno Diabetes Center, Aarhus University Hospital, Holstebro Hospital, and Hospital of South West Jutland, Esbjerg). The study was approved by the Ethics Committee of the Central Denmark Region (journal no. 20080229) and was conducted in accordance with the Helsinki Declaration. All participants provided written informed consent.

Cogen

The Copenhagen Cardiovascular Genetic study (COGEN) is a biobank that has collected superfluous whole blood from patients admitted to six cardiology departments in the greater region of Copenhagen from 2010-2017. COGEN currently contains samples from ~80,000 individuals. Due to a Danish permanent identification number (given to all permanent residents), these data can be linked with various clinical databases at an individual level. The present study population is comprised of individuals who have had at least one coronary angiogram performed between 2010-2014. From the Eastern Danish Heart Registry, which is a clinical database, we gathered information on demographics (e.g. age, sex, and BMI) and comorbidities (e.g. diabetes) which has been routinely entered on all patients in the Eastern region of Denmark. Information on e.g. glucose and HbA1c was gathered from routine blood samples.

Goya

Goya is a case-cohort study [7] where cases and equal numbers of population-based controls (n=2,740) were drawn from two large Danish cohorts. The two different cohorts used, were the Danish National Birth Cohort for women and a draft board examination cohort for men, respectively. Both were constituted by young Danish adults and the cases identified as those with most extreme BMI scores, whereas controls were randomly sampled. The male control group were identified from the records of 362,200 Caucasian men examined at the mean age of 20 years at the draft boards in Copenhagen and its surroundings during 1943–77. The female cases and control group were selected from 91,387 pregnant women recruited to the Danish National Birth Cohort during 1996–2002. In the present study, a total of 5373 genotyped individuals were included. Cases and controls were analyzed separately. The study was conducted as part of the activities of the Danish Obesity Research Center (DanORC, www.danorc.dk) and the MRC center for Causal Analyses in Translational Epidemiology (MRC CAiTE).

Health2006

The Health2006 cohort (KA20060011), is a population-based epidemiological study with the aim to assess the population prevalence of risk factors for different chronic diseases including, e.g., diabetes mellitus, and cardiovascular disease. It is composed as a random sample of the population aged 18 to 74 years and living in the south-western part of the greater Copenhagen area. A total of 3471 patients entered the study and participated in a health examination between June 2006 and June 2008 of which a subset was genotyped. In total, 2415 fully genotyped individuals were included in the present study**.** Written informed consent was obtained from all participants, and the study was approved by the Scientific Ethics Committee of the Capital Region of Denmark (KA-20060011) and was in accordance with the principles of the Declaration of Helsinki II. https://clinicaltrials.gov/ct2/show/record/NCT00316667.

Inter99

Inter99 is a population based intervention cohort [8], comprised of individuals from the Copenhagen area. Altogether 6161 individuals participated in the baseline examination, were genotyped and included in the present study. The Inter99 study is funded by The Danish Medical Research Council, The Danish Center for Evaluation and Health Technology Assessment, Novo Nordisk, Copenhagen County, The Danish Heart Foundation, The Danish Pharmaceutical Association, Augustinus foundation, Ib Henriksen foundation and Becket foundation.

Vejle Biobank

Vejle Biobank was established as a regional biobank between 2007 and 2010 [9]. The population, aged 25-75 years contains diabetes patients age- and gender -matched to a control population recruited from the Danish civil registry. Included patients were diagnosed if they had at least one glycated hemoglobin (HbA1c) value ≥6.6% (≥48.6 mmol/mol). 2,415 individuals were genotyped and included in this study.

EPIC-InterAct Study

The prospective InterAct type 2 diabetes case-cohort study [10] is coordinated by the MRC Epidemiology Unit in Cambridge and nested within the European Prospective Investigation into Cancer and Nutrition (EPIC). EPIC was initiated in the late 1980s and involves collaboration between 23 research institutions across Europe in 10 countries (Denmark, France, Germany, Greece, Italy, the Netherlands, Norway, Spain, Sweden and the United Kingdom). The majority of EPIC cohorts were recruited from the general population, with some exceptions. EPIC-InterAct included participants from 8 European countries, as data from Norway and Greece were not available. The case-cohort sample included 12,403 cases of type 2 diabetes and 16,154 subcohort members. Samples in EPIC Cambridge centre were excluded from this analysis because they overlap with EPIC-Norfolk study (see description below).

EPIC-InterAct samples were genotyped on Illumina 660w quad chip and Human Core-Exome-12v1-0_Bchip. Quality control was performed and failed samples were removed from further analysis. Data used in this analysis includes 8364 and 6745 participants genotyped on Illumina 660w quad chip and Human Core-Exome-12v1-0_Bchip, respectively. Analysis was done by chip.

EPIC-Norfolk

The EPIC-Norfolk study [11] is a prospective population-based cohort study which recruited 25,639 men and women aged 40-79 years at baseline between 1993 and 1997 from 35 participating general practices in Norfolk, United Kingdom. Individuals attended for a baseline health check including the provision of blood samples for concurrent and future analysis. They provided consent to future linkage to medical record information and a wide range of follow-up studies for different disease endpoints (including incident T2DM) have subsequently been undertaken , and further health check visits have been conducted since the baseline visit (see www.srl.cam.ac.uk/epic).

DNA has been extracted from all EPIC participants and stored blood has been analysed for an extensive range of classical and novel biomarkers. Sample quality control was performed including gender check, relatedness check, and ancestry check. Samples that failed quality control were removed from further analysis. Data used in this analysis includes 19,225 participants and their measures at the baseline health check.

Family Heart Study

The NHLBI FamHS study [12] design, collection of phenotypes and covariates as well as clinical examination have been previously described [1]. In brief, the FamHS recruited 1,200 families (approximately 6,000 individuals), half randomly sampled, and half selected for family history of coronary heart disease (CHD) or risk factor abnormalities as compared with age- and sex-specific population rates. The participants were sampled from four population-based parent studies: the Framingham Heart Study, the Utah Family Tree Study, and two centers for the Atherosclerosis Risk in Communities study (ARIC: Minneapolis, MN, and Forsyth County, NC). These individuals attended a clinic exam (1994-1996) where a broad range of phenotypes were assessed in the general domains of CHD, atherosclerosis, cardiac and vascular function, inflammation and hemostasis, lipids and lipoproteins, blood pressure, diabetes and insulin resistance, pulmonary function, diet, education, socioeconomic status, habitual behavior, physical activity, anthropometry, medical history and medication use. Approximately 8 years later, study participants belonging to the largest pedigrees were invited for a second clinical exam (2002-04). The most important CHD risk factors were measured again, including lipids, parameters of glucose metabolism, blood pressure, anthropometry, and several biochemical and hematologic markers. In addition, a computed tomography examination provided measures of coronary and aortic calcification, and abdominal and liver fat burden. Medical history and medication use was updated. A total of 2,756 European ancestry subjects in 510 extended random and high CHD risk families were studied. Also, 633 African ancestry subjects were recruited at ARIC field center at the University of Alabama in Birmingham. Informed consent was obtained from all participants.

From FamHS, 2,098 subjects were genotyped on Illumina Human Exome 12v1.0 BeadChip. Quality controls including Mendelian errors were assessed with LOKI, and pedigree relationship was checked using the pairwise correlation among exomechip autosome and GWAS autosome data. Originally 172 mitochondrial markers were genotyped. Based on CHARGEmtDNA+ WG probe blast, all these MT markers were valid and had the same position as in the Revised Cambridge Reference Sequence (rCRS) of the Human Mitochondrial DNA. There were 11 heterozygous mitochondrial haploid genotypes, which were set to missing before imputation via PLINK. We were able to impute 3631 MT markers with IMPUTE2 and the latest 1000 Human Genomes Reference (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/) released for MT in April, 2016. Based on the imputation accuracy assessment implemented within IMPUTE2, the concordance rate between original MT marker genotypes and the same ones imputed was 99.9%. After fcGENE software converted probabilities per cell to dosage and filtering, 390 markers remained as polymorphic and used for single variant association (Regression) and SKAT gene based association testing (seqMeta R package).

Fenland Study

The Fenland study [13] is a population-based cohort study that uses objective measures of disease exposure, such as accurate methods of body composition and energy expenditure, to study the

interactions between genetic and lifestyle factors that cause obesity and diabetes. 12,435 people were recruited from general practice lists in and around Cambridgeshire (Cambridge, Ely, and Wisbech) in the United Kingdom from birth cohorts from 1950–1975 during 2005-2015 in the first phase of the Fenland Study. 8,994 baseline samples were genotyped on Affymetrix Axiom UKBiobank chip, of which 8,453 samples passed quality control and were included in this analysis.

Framingham Heart Study

The Framingham Heart Study (FHS) [14,15] is a population-based, prospective study, which aims to examine the natural history, risk factors, and prognosis of cardiovascular, lung, and other diseases. The study began in 1948 with the recruitment of an original cohort of 5,209 men and women from the town of Framingham, Massachusetts. In 1971, the study recruited its offspring cohort, including 5,124 offspring and spouses of offspring of the FHS Original cohort. The Third Generation cohort started recruiting the adult children (n=4,095) of the offspring cohort participants in 2002. Participants underwent examinations every two years (the original cohort) or every four to eight years (the offspring and the third generation) to collect demographic and clinical measures and medical history .

A total of 199 unique mitochondrial DNA (mtDNA) variants in 8191 FHS individuals from two platforms were used for imputation. One platform was a customized array including 40 mtDNA variants genotyped by Illumina. The second platform was the human exome chip genotyping array by Illumina. All markers were aligned to correct positions according to the revised Cambridge Reference Sequence (rCRS). Prior to imputation, all markers were examined for incompatibilities in maternal lineages. Inconsistent alleles in pedigree members were set to missing. Any heteroplasmic alleles were also set to missing.

Mitochondrial DNA imputation was based on the Cosmopolitan 1000 Genomes Phase 3 mtDNA of 2505 individuals from 26 populations conforming with the rCRS of the Human mtDNA positions known also as v.5 phase 3 of 1000 Human Genomes HG19, GRCh37 cosmopolitan mitochondrial reference panel (ftp.1000genomes.ebi.ac.uk/vol1/ftp/) released for mitochondrial DNA in 2016. The imputation of mtDNA variants followed the CHARGEmtDNA+ working group (WG) protocol, using PLINK software for preparing FHS data in binary format, SHAPEIT2 for pre-phasing FHS mtDNA haplotype scaffold and IMPUTE2 for performing imputation of the full MT genome. The imputation quality was 99.9%, yielding 2636 mtDNA variants. After filtering in fcGENE, 639 markers remained for subsequent analyses.

This study includes individuals from the Offspring cohort and the Third Generation cohort. The proportion of female =53.6%.

BMI was measured at exam 7 for Offspring and exam 1 for Generation 3. Fasting glucose and insulin, HOMA-b and HOMA-IR were measured at exam 5 for Offspring and exam 1 for Generation 3. Hba1c was measured at exam 8 for Offspring. WHR was measured at exam 8 for Offspring and exam 2 for Generation 3. The exam 8 for Offspring and exam 2 for Generation 3 were at similar time periods.

GeneSTAR (Genetic Study of Atherosclerosis Risk)

GeneSTAR began in 1982 as the Johns Hopkins Sibling and Family Heart Study, a prospective longitudinal family-based study conducted originally in healthy adult siblings of people with

documented early onset coronary disease under 60 years of age who were hospitalized in one of ten Baltimore, Maryland hospitals. Siblings who were free of known coronary disease, autoimmune disease, and any life-threatening disease, and not pregnant, using chronic glucocorticosteroid therapy, or reporting history of chest radiation exposure were recruited. Between 1983 and 2007, 1656 European- and African-American siblings completed a baseline examination including demographics, medical history, medication use, physical examination by a cardiologist, exercise treadmill test using a modified Bruce protocol, family history, diet, anthropometrics, blood pressure measurement, smoking status, physical activity and fasting blood measures including lipids, glucose, insulin, fibrinogen, Lp(a), and inflammatory markers [16-18], Siblings are followed regularly via questionnaire or repeat screening for incident cardiovascular disease, stroke, peripheral arterial disease, diabetes, cancer, and related comorbidities, from 10 to 35 years after study entry [19,20]. Commencing in 2003, the siblings, their offspring, and the coparent of the offspring (total N=3003) participated in a 2 week trial of aspirin 81 mg/day with pre and post ex vivo platelet function assessed using multiple agonists in whole blood and platelet rich plasma along with assessment of anthropometrics, blood pressure measurement, smoking status, physical activity and fasting blood measures including lipids, glucose, fibrinogen, and inflammatory markers [21,22]. The offspring are also followed regularly using the same protocol and outcomes as the siblings. The GeneSTAR study was approved by the Johns Hopkins Medicine Institutional Review Board, and all participants gave written informed consent prior to screening.

Genotyping was performed with the Illumina 1M_v1C chip in 3,232 GeneSTAR participants at deCODE Genetics in Reykjavik, Iceland. Standard quality control measures were employed, including duplicates and controls on each plate, gender checks, and Mendelian consistency checks. All samples with gender discrepancies, >10% missing data, >5% Mendelian inconsistencies, or outliers from EIGENSTRAT principal component analysis were removed prior to analysis or imputation. There were 162 mitochondrial markers genotyped. Positions were remapped using the Revised Cambridge Reference Sequence of the Human Mitochondrial DNA per the CHARGE mtDNA working group. No heterozygous haploid genotypes or SNPs with 10% missing were identified. We were able to impute 3642 MT markers in both European Americans and African Americans separately using IMPUTE2 and the latest 1000 Human Genomes Reference (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/) released for MT in April, 2016. Based on the imputation accuracy assessment implemented within IMPUTE2, the concordance rate between original MT marker genotypes and the same ones imputed was 99.8% for both European Americans and African Americans. After fcGENE software converted probabilities per cell to dosage and filtering, 492 markers for European Americans and 714 markers for African Americans remained as polymorphic and were used for single variant association (regression) and SKAT gene based association testing (seqMeta R package).

GENOA (Genetic Epidemiology Network of Arteriopathy)

The GENOA cohorts recruited non-Hispanic African American (AA) and white participants (EA) from two centers between 1996 and 2000 [23]. As described previously, in Jackson, MS, USA, residents with essential hypertension were identified through the Atherosclerosis Risk in Communities cohort, a general population sample of 45- to 64-year-old non-Hispanic African American residents. The non-Hispanic white subjects in the present study were members of sibships initially enrolled in Rochester, MN, USA between July *1997* and August 1999 in the

GENOA of the Family Blood Pressure Program (FBPP) [24]. The FBPP, sponsored by the National Heart, Lung, and Blood Institute, was designed to identify and characterize genetic determinants of hypertension and its associated cardiac and renal complications. For the GENOA–Rochester cohort, the Mayo Clinic diagnostic index and medical record linkage system of the Rochester Epidemiology Project [25] were used to identify non-Hispanic white residents of Olmsted County with a diagnosis of essential hypertension made before age 60. Probands with evident secondary hypertension or advanced chronic kidney disease (serum creatinine >2.0 mg/dl) were not recruited at either site. The hypertensive proband and all siblings were invited to participate if at least one other sibling had essential hypertension. Between 2000 and 2004, 2721 (79%) of the 3434 original GENOA participants (1482 African Americans and 1239 whites) returned for a second clinic visit. All subjects provided consent and all protocols were approved by the Mayo Clinic and University of Mississippi Medical Center Institutional Review Boards.

In the African American cohort, subjects were genotyped primarily on the Affymetrix SNP Array 6.0. Samples that failed on the Affymetrix SNP Array 6.0 were run again on the Illumina 1M-Duo platform. Then, the genotype data for the African American cohort were merged. In the non-Hispanic white cohort, subjects were genotyped primarily on the Affymetrix SNP Array 6.0. Samples that failed on the Affymetrix SNP Array 6.0 were run again on the Illumina 660W and Illumina 1M-Duo platforms. Then, the genotype data for the non-Hispanic white cohort were merged.

Prior to mitochondrial (MT) marker imputation, 90 MT markers were available in the AA cohort genotype data and 56 in the EA cohort, respectively. Further, based on CHARGEmtDNA+ WG probe blast, 63 MT markers in the AA and 45 markers in the EA subjects were found valid and were updated with the latest Revised Cambridge Reference Sequence (rCRS) positions of the Human Mitochondrial DNA. The sets of 63 and 45 MT markers were used separately as scaffold for haplotype prephasing and for MT imputation.

Mitochondrial imputation used the latest 1000 Human Genomes Reference (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/) released for MT in April 2016. The procedure of MT imputation followed the CHARGEmtDNA+ working group (WG) protocol, using PLINK software for preparing the data in binary format, SHAPEIT2 for prephasing MT haplotype and IMPUTE2 for performing imputation of the full MT genome, separately for AA and EA cohorts.

The GENOA imputed dosage file (after fcGENE converting and filtering) contained 3,617 good quality MT markers for 1,263 AA subjects and 3,615 good quality MT markers for 1,386 EA subjects, respectively. Based on the IMPUTE2 imputation accuracy assessment, the imputation concordance rate between original MT genotypes and the same imputed ones was 99.5% for AA cohort and 99.3% for EA cohort, respectively. Among the good quality imputed MT markers, only 334 MT markers had R^2>=0.3 and were polymorphic in AA cohort and 134 MT markers in EA cohort, respectively. A few heterozygotes in the imputed data were considered as possible genotyping/imputation errors and turned to missing prior to association analysis.

GOLDN (Genetics of Lipid Lowering Drugs and Diet Network Study)

GOLDN participants were recruited from 3-generation families previously screened in the NHLBI Family Heart Study (FHS) Minnesota or Utah centers [12]. Because FHS was conducted mostly in participants of Caucasian descent, the GOLDN population is predominantly white. Excluded were people with fasting TGs ≥1500 mg/d; recent history of myocardial infarction or coronary revascularization; a history of liver, kidney, pancreas, or gall bladder disease; a history of nutrient malabsorption; current use of insulin; abnormal liver or kidney function; or currently pregnant or nursing. Individuals who currently used prescription or dietary supplements that influenced lipids were required to discontinue them for 4 weeks prior to study. The GOLDN sample consisted of 1048 individuals in 184 pedigrees (average family size 6.2); mean age at intervention was 49±16, 48% male, 51% from MN, and 8% were current smokers. This gene-environment interaction study employed two interventions, one to raise lipids and one to lower lipids. In the post-prandial lipemia (PPL) intervention, participants fasted for ≥12 hr and abstained from alcohol intake for ≥24. The PPL intervention followed the protocol of Patsch *et al*. (1992) [26]. The whipping cream (83% fat) meal had 700 Calories/m^2 body surface area; 3% of calories were derived from protein, and 14% from carbohydrate. Blood samples were drawn immediately before (fasting) and at 3.5 and 6 hours after consuming the high-fat meal. During the 6-hour study period, participants consumed only water and abstained from physical activity. The lipid-lowering intervention was a 3-wk, open-label trial of fenofibrate (FFB, 160 mg qd). A detailed description of the GOLDN study design can be found in Lai, *et al*. (2007) [27].

Out of 1048 individuals in the study, only 862 subjects were genotyped on Affy6.0. Originally, 110 mitochondrial markers were genotyped in Affy6.0. Based on CHARGEmtDNA+ WG probe blast, all these MT markers were valid and had the same position as in the Revised Cambridge Reference Sequence (rCRS) of the Human Mitochondrial DNA. We were able to impute 3603 MT markers with IMPUTE2 and the latest 1000 Human Genomes Reference (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/) released for MT in April 2016. Based on the imputation accuracy assessment implemented within IMPUTE2, the concordance rate between original MT marker genotypes and the same ones imputed was 100%. After R^2 > 0.3 filtering, 3129 markers remained and were used for single variant association (Regression) and SKAT gene based association testing (seqMeta R package).

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL)

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a community-based cohort study of self-identified Hispanic/Latino individuals from four US metropolitan areas [28]. As described previously, the HCHS/SOL sample survey design consisted of a two-stage probability sample of households at each of four recruitment centers: Chicago, Miami, the Bronx, and San Diego [29]. Census block groups were selected in defined communities near each center, and households were sampled within block groups. Households with Hispanic/Latino surnames and individuals were oversampled as a means of increasing representation of the Hispanic/Latino target population; likewise, households with residents over 45 years of age were oversampled so a more uniform age distribution could be achieved. Sampling weights were calculated for each individual to reflect the probability of sampling. Baseline examination methods were described by Sorlie *et al* [28]. The HCHS/SOL study was approved by institutional review boards at participating institutions, and written informed consent was obtained from all participants.

DNA extracted from blood was genotyped on an Illumina custom array, SOL HCHS Custom 15041502 B3, consisting of the Illumina Omni 2.5M array (HumanOmni2.5-8v1-1) and ~150,000 custom SNPs selected to include ancestry-informative markers, variants characteristic of Amerindian populations, previously identified GWAS hits, and other candidate-gene polymorphisms. Samples were checked for annotated sex or genetically determined sex, gross chromosomal anomalies [30], unexpected duplicates, missing call rates, contamination, and batch effects [31]. Portions of the genome with large chromosomal anomalies were filtered out in 71 samples. A total of 12,803 samples passed quality control with a missing call rate < 1%. Quality metrics used to filter SNPs for the imputation basis and association testing included missing call rate (>2%), Mendelian errors (>3 in 1,343 trios or duos), duplicate-sample discordance (>2 in 291 sample pairs), and deviation from Hardy-Weinberg equilibrium ($p < 10^{-5}$ in a meta-analysis of nine groups within which individuals had both parents from the same country of origin). SNPs were regarded as "informative" if they had no positional duplicate on the array and were polymorphic in the sample. A total of 2,232,944 SNPs passed quality metrics and were informative. Genotype imputation was performed with the 1000 Genomes Project phase 1 reference panel [32]. The 12,803 samples were imputed together with genotyped SNPs that passed quality filters and represented unique positions on the autosomes and non-pseudo-autosomal parts of the X chromosome. For the current project, we updated the list of individuals providing consent,as of 2016, which included 12,434 individuals. After eliminating SNPs that failed quality metrics, 321 high quality mtDNA variants were available for inclusion in these analyses. Among these 321 variants, 209 had a minor allele frequency (MAF) of at least 1%-2%. The imputation of mtDNA variants followed the CHARGEmtDNA+ working group (WG) protocol. We used PLINK software to prepare data in binary format, SHAPEIT2 for pre-phasing mtDNA haplotypes and IMPUTE2 combined with the latest 1000 Human Genomes Reference (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/), April 2016 release, for performing imputation of the full mitochondrial genome. The imputation quality in 99.3%, yielding 3667 mtDNA variants. After applying fcGENE to convert probabilities per cell to dosage, 757 mtDNA variants remained polymorphic.

HyperGEN

HyperGEN is a family-based study that looks at the genetic causes of hypertension and related conditions in EA and AA subjects. HyperGEN recruited hypertensive sibships, along with their normotensive adult offspring, and an age-matched random sample. Families were drawn from population-based cohorts or the community-at-large if sibships had ≥2 siblings who had been diagnosed with hypertension before age 60. The study was later extended to include siblings and offspring of the original sibpairs. Hypertension was defined as current antihypertensive medication use or having an average systolic blood pressure ≥140 mm Hg and/or diastolic blood pressure ≥90 mm Hg measured at two clinic visits. In total, HyperGEN has collected data on 2,471 Caucasian-American subjects (2,029 in 455 extended families) and 2,300 African-American subjects (1,886 from 584 families), from five field centers in Alabama, Massachusetts, Minnesota, North Carolina, and Utah. The HyperGEN study design and methods have been previously described [33,34]. In the present report, we used genetic marker data from 1,695 CA and

1,258 AA subjects genotyped using Illumina Cardio-MetaboChip and Affymetrix SNP Array 6.0, respectively.

In the HyperGEN AA cohort, after applying established QC procedures we retained data of 1,258 subjects genotyped primarily on the Affymetrix SNP Array 6.0. A total of 44 mitochondrial markers from the Affymetrix SNP Array 6.0 data remained in the AA cohort after initial QC. Similarly in the HyperGEN CA cohort, after QC we included data of 1,695 subjects genotyped on the Illumina Cardio-MetaboChip. The ChargeMT DNA working group protocol was applied to further process the MT genotype data. Based on CHARGEmtDNA+ WG probe blast, 44 markers in the AA cohort and 52 in the CA cohort were valid and had the same position as in the Revised Cambridge Reference Sequence (rCRS) of the Human Mitochondrial DNA. These markers were further used for imputation and association analysis. The imputation of mtDNA variants followed the CHARGEmtDNA+ working group (WG) protocol, using PLINK software for preparing data in binary format, SHAPEIT2 for pre-phasing mtDNA haplotype scaffold and IMPUTE2 combined with the latest 1000 Human Genomes Reference (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/), April 2016 release, for performing imputation of the full MT genome. The imputation quality in AA was 86.2%, yielding 3616 mtDNA variants; whereas in CA the imputation quality was 98.4%, yielding 3618 mtDNA variants. After applying fcGENE to convert probabilities per cell to dosage, 118 and 119 MT markers remained polymorphic, in AA and CA cohorts, respectively, and were used in a regression based single-variant association analysis. The R package seqMeta was used to derive gene-based statistics for association testing.

Long Life Family Study (LLFS)

LLFS is a family-based cohort study, including four clinical centers: Boston University Medical Center in Boston, MA, USA, Columbia University Medical School, New York, NY, USA**,** University of Pittsburgh in Pittsburgh PA, USA, and University of Southern Denmark, Denmark. The study characteristics, recruitment, eligibility and enrollment have been previously described [35-37]. In brief, the LLFS was designed to determine genetic, behavioral, and environmental factors related to families of exceptionally healthy, elderly individuals. Phase 1 was conducted between 2006 and 2009 recruiting 4,953 individuals from 539 families. The probands were at least 79 years old in the USA centers, and 90 years old or older in Denmark. The families were selected to participate in the study based on The Family Longevity Selection Score (FLoSS) [36], a score generated according to birth-year cohort survival probabilities of the proband and their siblings. Probands and their families with FLoSS score of 7 or higher, at least one living sibling, and at least one living offspring (minimum family size of 3), who were able to give informed consent and willing to participate were recruited. The individuals were genotyped using ~2.5 million SNPs from the Illumina Omni chip. The LLFS is currently gathering data during a second in-home visit, averaging about 7-8 years after the first visit.

Prior to mitochondrial (MT) marker imputation, 253 MT markers were examined for typing incompatibilities in pedigrees, and 32 discrepant genotypes in 30 pedigrees were identified. The corresponding non-conform genotypes were zeroed out for a whole matrilineage group for an erroneous marker. Further, based on work of CHARGEmtDNA+ WG collaboration, blast probe

matches of LLFS MT markers of Illumina Omni chip, 110 MT markers were found not valid and dropped from the analysis. The remaining 143 MT markers were updated with latest MT Cambridge Revised Sequence (rCRS) of the Human MT-DNA positions, and used as scaffold for haplotype prephasing and for MT imputation.

Mitochondrial imputation were based on the Cosmopolitan 1000 Genomes Phase 3 MT DNA of 2505 individuals from 26 populations conforming with the rCRS of the Human MT-DNA positions known also as v.5 phase 3 of 1000 Human Genomes HG19, GRCh37 cosmopolitan mitochondrial reference panel (ftp.1000genomes.ebi.ac.uk/vol1/ftp/) released for MT in 2016. The procedure of MT imputation followed the CHARGEmtDNA+ working group (WG) protocol, using PLINK software for preparing LLFS data in binary format, SHAPEIT2 for prephasing LLFS MT haplotype scaffold and IMPUTE2 for performing LLFS imputation in one window in the length of the full MT genome.

The final LLFS imputed dosage file (after fcGENE converting and filtering) contained 451 polymorphic MT markers for 4,710 subjects. Based on the IMPUTE2 imputation accuracy assessment, the imputation concordance rate between original MT genotypes and the same imputed ones for LLFS was 99.7%. A few heterozygotes in the original data were considered as possible genotyping errors and turned to missing via PLINK.

Lothian Birth Cohort of 1921 (LBC1921)

The LBC1921[38-40] consists of 550 participants (316 female) who were born in 1921 and at age 11 took part in the Scottish Mental survey of 1932. Individuals who took part in the Scottish Mental Survey were identified through examining the records of those registered with a general practitioner in the area. At age 79 these individuals were followed up and those living in Edinburgh and the surrounding regions were recruited into the LBC1921 cohort. They are healthy older age individuals living independently within the community. Both BMI and Hba1c were measured at age 79. Venous whole blood was extracted following informed consent and ethical approval was given from The Lothian Research Ethics Committee.

Lothian Birth Cohort 1936 (LBC1936)

The LBC1936[41] has a total of 1091 participants (543 female) who took part in the Scottish Mental Survey of 1947. These individuals were recruited in a similar fashion to the participants of the LBC1921 cohort. At age 70 these individuals were recruited into the LBC1936 where both BMI and Hba1c were measured. These participants were healthy individuals who were able to live independently within the community. Venous whole blood was extracted following informed consent and ethical approval was granted by Scotland's Multi-Center Research Ethics Committee and the Lothian Research Ethics Committee.

MESA (Multi-Ethnic Study of Atherosclerosis)

The Multi-Ethnic Study of Atherosclerosis (MESA) is a study of the characteristics of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. MESA consisted of a diverse, population-based sample of an initial 6,814 asymptomatic men and women aged 45-84. 38 percent of the recruited participants were white, 28 percent African American, 22 percent Hispanic, and 12 percent Asian, predominantly of Chinese descent. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern

University and University of California - Los Angeles. This study was approved by the IRB of each study site, and written informed consent was obtained from all participants. Each participant received an extensive physical exam and determination of coronary calcification, ventricular mass and function, flow-mediated endothelial vasodilation, carotid intimal-medial wall thickness and presence of echogenic lucencies in the carotid artery, lower extremity vascular insufficiency, arterial wave forms, electrocardiographic (ECG) measures, standard coronary risk factors, sociodemographic factors, lifestyle factors, and psychosocial factors. Selected repetition of subclinical disease measures and risk factors at follow-up visits allowed study of the progression of disease. Participants are being followed for identification and characterization of cardiovascular disease events, including acute myocardial infarction and other forms of coronary heart disease (CHD), stroke, and congestive heart failure; for cardiovascular disease interventions; and for mortality. The first examination took place over two years, from July 2000 - July 2002. It was followed by four examination periods that were 17-20 months in length. Participants have been contacted every 9 to 12 months throughout the study to assess clinical morbidity and mortality.

NEO (The Netherlands Epidemiology of Obesity study)

The NEO was designed for extensive phenotyping to investigate pathways that lead to obesity-related diseases. The NEO study is a population-based, prospective cohort study that includes 6,671 individuals aged 45–65 years, with an oversampling of individuals with overweight or obesity. At baseline, information on demography, lifestyle, and medical history have been collected by questionnaires. In addition, samples of 24-h urine, fasting and postprandial blood plasma and serum, and DNA were collected. Genotyping was performed using the Illumina HumanCoreExome chip, which was subsequently imputed to the 1000 genome reference panel. Participants underwent an extensive physical examination, including anthropometry, electrocardiography, spirometry, and measurement of the carotid artery intima-media thickness by ultrasonography. In random subsamples of participants, magnetic resonance imaging of abdominal fat, pulse wave velocity of the aorta, heart, and brain, magnetic resonance spectroscopy of the liver, indirect calorimetry, dual energy X-ray absorptiometry, or accelerometry measurements were performed. The collection of data started in September 2008 and completed at the end of September 2012. Participants are currently being followed for the incidence of obesity-related diseases and mortality.

OOA (The Old Order Amish Study)

The Old Order Amish of Lancaster County, PA are a genetically homogeneous population whose ancestors emigrated to the Lancaster area in the 1700's from north central Europe. The current day population of ~35,000 Older Amish in this community are nearly all descendants of the initial immigrants. Since 1993, investigators at the University of Maryland have worked closely with the Amish community to study the genetics of a number of complex diseases and traits. These studies generally employ community-based recruitment of relatively healthy individuals and their family members [42-46]. All studies were approved by the Institutional Review Board of the University of Maryland, and all subjects provided written informed consent.

This report is based on 2,312 Amish study participants who were genotyped on the Infinium HumanCoreExome-24 BeadChip. Mitochondrial DNA imputation was based on the Cosmopolitan 1000 Genomes Phase 3 mtDNA of 2505 individuals from 26 populations conforming with the rCRS of the Human mtDNA positions known also as v.5 phase 3 of 1000 Human Genomes HG19,

GRCh37 cosmopolitan mitochondrial reference panel (ftp.1000genomes.ebi.ac.uk/vol1/ftp/) released for mitochondrial DNA in 2016. The imputation of mtDNA variants followed the CHARGEmtDNA+ working group (WG) protocol, using PLINK software for preparing LLFS data in binary format, SHAPEIT2 for pre-phasing FHS mtDNA haplotype scaffold and IMPUTE2 for performing imputation of the full MT genome. The imputation quality was 98.9%, yielding 3623 mtDNA variants. After filtering in fcGENE, 3623 markers remained for subsequent analyses.

The community has maintained detailed and accurate genealogical records of all members, and through these records, we are able to establish how each individual in the community is related to every other individual. To account for these relationships, association analysis of these phenotypes was performed under a variance component model that assesses the effect of genotype, as an additive effect, on the quantitative trait, while simultaneously estimating the effects of age, age^2, sex, and a polygenic component to account for phenotypic correlation due to relatedness. The polygenic component was modelled using the relationship matrix derived from the pedigree structure to account for the relatedness of all subjects in the study. Genome-wide analysis using the a smaller version of the pedigree was carried out using the R software package.

1982 Pelotas (Brazil) Birth Cohort Study

The maternity hospitals in Pelotas, a southern Brazilian city (current population ~330,000), were visited daily in the year of 1982. The 5,914 liveborns whose families lived in the urban area were examined and their mothers interviewed. Information was obtained for more than 99% of the livebirths. These subjects have been followed-up at the following mean ages: 11.3 months (all children born from January to April 1982; n=1457), 19.4 months (entire cohort; n=4934), 43.1 months (entire cohort; n=4742), 13.1 years (random subsample; n=715), 14.7 years (systematic subsample; n=1076); 18.2 (male cohorts attending to compulsory Army recruitment examination; n=2250), 18.9 (systematic subsample; n=1031), 22.8 years (entire cohort; n=4297) and 30.2 years (entire cohort; n=3701). Details about follow-up visits and available data can be found in the two Cohort Profile papers (PMID: 16373375 and 25733577). DNA samples (collected at the mean age of 22.8 years) were genotyped for ~2.5 million of SNPs using the Illumina HumanOmni2.5-8v1 array (which includes autosomal, X and Y chromosomes, and mitochondrial variants). After quality control, the data were prephased using SHAPEIT and imputed using IMPUTE2 based on 1000 Genomes haplotypes.

ROTTERDAM STUDY (II and III)

Rotterdam Study (RS) is a population-based cohort study in Rotterdam, the Netherlands. The design of the RS has been previously described in detail elsewhere [47]. In brief, the RS includes three sub-cohorts. In 1990, all residents of Ommoord, a district in Rotterdam, aged 45 years and older were invited to participate (RS-I). In 2000, the cohort was extended with 3,011 participants who reached the age 55 years or who were 55 years and over and had moved into the research area (RS-II). In 2006, a third cohort of 3,934 participant aged 45 years and older was initiated (RS-III). Demographic and clinical variables were collected via standardized questionnaires at each visit and peripheral blood was collected for later analysis. The study was approved by the medical ethics committee at Erasmus University Rotterdam, Rotterdam, the Netherlands, and all

examined participants gave written informed consent. The RS participants were genotyped using the Illumina Infinium HumanHap550K v.3 (RS-I and RS-II) and HapMap 610 (RS-III) in the Genetic Laboratory of Erasmus MC Department of Internal Medicine, the Netherlands, following manufacturers' protocols and quality control standards.

A total of 199 unique mitochondrial DNA (mtDNA) variants from two platforms were used for imputation. One platform was a 550K v.3 including 40 mtDNA variants genotyped by Illumina. The second platform was 610 by Illumina. Further, based on work of CHARGEmtDNA+ WG collaboration, blast probe matches of RS MT markers of Illumina Omni chip, 100 MT markers were found not valid and dropped from the analysis. The remaining 143 MT markers were updated with latest MT Cambridge Revised Sequence (rCRS) of the Human MT-DNA positions, and used as scaffold for haplotype prephasing and for MT imputation.

Mitochondrial imputation were based on the Cosmopolitan 1000 Genomes Phase 3 MT DNA of 2504 individuals from 26 populations conforming with the rCRS of the Human MT-DNA positions known also as v.5 phase 3 of 1000 Human Genomes HG19, GRCh37 cosmopolitan mitochondrial reference panel (ftp.1000genomes.ebi.ac.uk/vol1/ftp/) released for MT in 2016. The procedure of MT imputation followed the CHARGEmtDNA+ working group (WG) protocol, using PLINK software for preparing RS data in binary format, SHAPEIT2 for prephasing LLFS MT haplotype scaffold and IMPUTE2 for performing RS imputation in one window in the length of the full MT genome.

The final RS imputed dosage file (after fcGENE converting and filtering) contained 451 polymorphic MT markers. Based on the IMPUTE2 imputation accuracy assessment, the imputation concordance rate between original MT genotypes and the same imputed ones for RS was 99.7%. A few heterozygotes in the original data were considered as possible genotyping errors and turned to missing via PLINK.

SHIP (Study of Health In Pomerania)

The Study of Health In Pomerania [48] is a prospective longitudinal population-based cohort study in Western Pomerania assessing the prevalence and incidence of common diseases and their risk factors. SHIP encompasses the two independent cohorts SHIP and SHIP-TREND. Participants aged 20 to 79 with German citizenship and principal residency in the study area were recruited from a random sample of residents living in the three local cities, 12 towns as well as 17 randomly selected smaller towns. Individuals were randomly selected stratified by age and sex in proportion to population size of the city, town or small towns, respectively. A total of 4,308 participants were recruited between 1997 and 2001 in the SHIP cohort. Between 2008 and 2012 a total of 4,420 participants were recruited in the SHIP-TREND cohort. Individuals were invited to the SHIP study center for a computer-assisted personal interviews and extensive physical examinations. Blood samples were taken between 07:00 AM and 04:00 PM, and serum aliquots were prepared for immediate analysis and for storage at -80 °C in the Integrated Research Biobank (Liconic, Liechtenstein). Study participants were genotyped using the Illumina Infinium Human Exome BeadChip v1.0.

TUDR (Taiwan-US Diabetic Retinopathy Study)

Taiwan-US Diabetic Retinopathy Study (TUDR) 2009 to present, is a cohort that enrolled subjects with Type 2 diabetes receiving care at Taichung Veteran General Hospital (Taichung VGH), and a small number of subjects from Taipei Tri-Service General Hospital. All TUDR subjects underwent a complete ophthalmic and fundus examination to carefully document the presence and extent of retinopathy.

WHI (The Women's Health Initiative)

WHI is a long-term, prospective, multi-center cohort study investigating post-menopausal women's health in the US. WHI was funded by the National Institutes of Health and the National Heart, Lung, and Blood Institute to study strategies to prevent heart disease, breast cancer, colon cancer, and osteoporotic fractures in women 50-79 years of age. WHI involves 161,808 women recruited between 1993 and 1998 at 40 centers across the US. The study consists of two parts: the WHI Clinical Trial which was a randomized clinical trial of hormone therapy, dietary modification, and calcium/Vitamin D supplementation, and the WHI Observational Study, which focused on many of the inequities in women's health research and provided practical information about incidence, risk factors, and interventions related to heart disease, cancer, and osteoporotic fractures. For this project, we used women who self-identified as African or Hispanic and were part of the WHI SHARe (SNP Health Association Resource) study sample https://www.nhlbi.nih.gov/whi . The dataset includes extensive phenotypic and genotypic data on 12008 African-American and Hispanic women. After exclusions for those missing phenotype and genotype data, a maximum of 8116 African and 3463 Hispanic ancestry women were included in analyses.

Genotyping was done on the Affymetrix 6.0 platform. DNA was extracted from blood specimens collected at time of WHI enrollment. All samples, plus 2% blinded duplicates, were genotyped at Affymetrix on the Genome-wide Human SNP Array 6.0 (909,622 SNPs). Approximately 1% of samples failed genotyping; samples with call rate less than 95%, unexpected duplicates, and samples with genotype calls on the Y chromosome were further excluded. A total of 8421 African-American and 3587 Hispanic women were genotyped. A total of 119 variants on the mitochondria were available for imputation for Hispanic and African ancestry individuals. The positions of these markers was updated using the latest MT Cambridge Revised Sequence (rCRS) for Human MT-DNA in preparation for the CHARGEmtDNA+ working group imputation protocol. We used PLINK software to prepare data in binary format, SHAPEIT2 for pre-phasing mtDNA haplotypes and IMPUTE2 combined with the latest 1000 Human Genomes Reference (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/), April 2016 release, for performing imputation of the full mitochondrial genome. In AA, the imputation quality was 91.7%, yielding 3678 mtDNA variants. In HA, imputation quality was 91.0%, yielding 3678 mtDNA variants. After applying fcGENE to convert probabilities per cell to dosage, 525 mtDNA variants in AA and 426 mtDNA variants in HA remained polymorphic.

Women's Genome Health Study (WGHS)

The Women's Genome Health Study (WGHS) is a prospective cohort of initially healthy, female North American health care professionals at least 45 years old at baseline representing participants in the Women's Health Study (WHS) who provided a blood sample at baseline and

consent for blood-based analyses. The WHS was a 2x2 trial beginning in 1992-1994 of vitamin E and low dose aspirin in prevention of cancer and cardiovascular disease with about 10 years of follow-up. Since the end of the trial, follow-up has continued in observational mode. Additional information related to health and lifestyle were collected by questionnaire throughout the WHS trial and continuing observational follow-up [49].

Genotype information for mitochondria derives from the Illumina HumanExome Beadchip v.1.1 among 22,618 WGHS participants with self-reported European ancestry that was verified by genetic analysis. Genotyped samples were initially retained with >98% successful genotyping of the SNP content. Samples and SNPs were then filtered according to the best practices protocol developed by the CHARGE consortium as reported in Grove *et al*. [50] within GenomeStudio v. 2011.1 software. This processing incorporated data-driven definitions of SNP clusters. The results of genotype calling in GenomeStudio were output in the report format that included scaled XY experimental intensity values as well as genotype calls for input into zCall (3, version. May 8, 2012). zCall [51] was run to identify exclusively heterozygous genotypes deemed to have acceptable quality but excluded by cluster definitions in GenomeStudio among SNPs that were retained by the best practices protocol. Homozygotes called by zCall were not used to replace missing genotypes from GenomeStudio; and no non-missing genotypes from GenomeStudio were replaced by genotypes identified by zCall. Subsequently, an additional 1,429 SNPs were excluded with a manually-validated statistical model that used the following parameters: cluster separation in GenomeStudio, the GenomeStudio R value for the major homozygote genotypes, the minor allele frequency from GenomeStudio, and the number of missing genotypes that remained not called by zCall. The final WGHS exome chip data include the total of 22,618 WGHS participants with verified European ancestry and successful genotype for 235,667 SNPs, of which 177,812 had minor allele frequency > 0. Imputation of genotype information that was not measured directly with the Exome chip was performed as described for the CHARGE mtDNA consortium.

WU-HHD (Washington University Hypertensive Heart Disease Study)

The Washington University Hypertensive Heart Disease (**WU-HHD**) study enrolled non-Hispanic Caucasian subjects (50% with Hypertension) to study the genetic determinants of hypertensive heart disease (HHD). Hypertensive and normotensive subjects representing a wide-spectrum of health traits were enrolled from 2002 to 2011. Along with demographic and anthropomorphic data, HHD and related phenotypes have been collected including ventricular and atrial structure, cardiovascular and metabolic health. The study design and methods have been previously described [52]. The Washington University IRB approved this study and all subjects provided written informed consent.

For the present report, we included 877 subjects enrolled and genotyped in the WU-HHD study. Genotyping was obtained by the Washington University Genome Technology Access Center (GTAC) using Affymetrix SNP Array 6.0. After genotype data QC using established procedures, we excluded related samples and poor quality arrays and retained 711 unrelated subjects for analysis. A total of 167 mitochondrial markers were initially collected from the Affymetrix SNP Array 6.0 data. The ChargeMT DNA working group protocol was applied to further process the

MT genotype data. Based on CHARGEmtDNA+ WG probe blast, 95 of the 167 markers were valid and had the same position as in the Revised Cambridge Reference Sequence (rCRS) of the Human Mitochondrial DNA. These markers were further used for imputation and association analysis. Following protocols developed by the CHARGEmtDNA+ working group, we first used PLINK to prepare the genotype data in binary format, and then applied shapeit2 to prophase and IMPUTE2 to finally impute the full MT genomes. A total of 3642 MT markers were imputed by IMPUTE2 using the latest 1000 Human Genomes Reference (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/), April 2016 release. The imputation concordance rate between original MT genotypes and the same imputed ones was 95.4%. After applying fcGENE to convert probabilities per cell to dosage, 190 markers remained polymorphic and were used in a regression based single-variant association analysis. The R package seqMeta was used to derive gene-based statistics for association testing.

References

1. Elliott, P. *et al.* The Airwave Health Monitoring Study of police officers and staff in Great Britain: rationale, design and methods. *Environ Res* **134**, 280-5 (2014).
2. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* **42**, 111-27 (2013).
3. Fraser, A. *et al.* Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* **42**, 97-110 (2013).
4. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**, 816-34 (2010).
5. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
6. Drivsholm, T., Ibsen, H., Schroll, M., Davidsen, M. & Borch-Johnsen, K. Increasing prevalence of diabetes mellitus and impaired glucose tolerance among 60-year-old Danes. *Diabet Med* **18**, 126-32 (2001).
7. Paternoster, L. *et al.* Genome-wide population-based association study of extremely overweight young adults--the GOYA study. *PLoS One* **6**, e24303 (2011).
8. Jorgensen, T. *et al.* A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99. *Eur J Cardiovasc Prev Rehabil* **10**, 377-86 (2003).
9. Petersen, E.R. *et al.* Vejle Diabetes Biobank - a resource for studies of the etiologies of diabetes and its comorbidities. *Clin Epidemiol* **8**, 393-413 (2016).
10. InterAct, C. *et al.* Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia* **54**, 2272-82 (2011).
11. Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer* **80 Suppl 1**, 95-103 (1999).
12. Higgins, M. *et al.* NHLBI Family Heart Study: objectives and design. *Am J Epidemiol* **143**, 1219-28 (1996).
13. Rolfe Ede, L. *et al.* Association between birth weight and visceral fat in adults. *Am J Clin Nutr* **92**, 347-52 (2010).

14. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M. & Castelli, W.P. The Framingham Offspring Study. Design and preliminary data. *Prev Med* **4**, 518-25 (1975).
15. Splansky, G.L. *et al.* The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* **165**, 1328-35 (2007).
16. Becker, D.M. & Levine, D.M. Risk perception, knowledge, and lifestyles in siblings of people with premature coronary disease. *Am J Prev Med* **3**, 45-50 (1987).
17. Becker, L.C. *et al.* Screening of asymptomatic siblings of patients with premature coronary artery disease. *Circulation* **75**, II14-7 (1987).
18. Becker, D.M. *et al.* Risk factors in siblings of people with premature coronary heart disease. *J Am Coll Cardiol* **12**, 1273-80 (1988).
19. Blumenthal, R.S. *et al.* Exercise thallium tomography predicts future clinically manifest coronary heart disease in a high-risk asymptomatic population. *Circulation* **93**, 915-23 (1996).
20. Vaidya, D. *et al.* Incidence of coronary artery disease in siblings of patients with premature coronary artery disease: 10 years of follow-up. *Am J Cardiol* **100**, 1410-5 (2007).
21. Becker, D.M. *et al.* Sex differences in platelet reactivity and response to low-dose aspirin therapy. *JAMA* **295**, 1420-7 (2006).
22. Mathias, R.A. *et al.* A combined genome-wide linkage and association approach to find susceptibility loci for platelet function phenotypes in European American and African American families with coronary artery disease. *BMC Med Genomics* **3**, 22 (2010).
23. Daniels, P.R. *et al.* Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *Am J Med* **116**, 676-81 (2004).
24. Morrison, A.C. *et al.* Genome scan for hypertension in nonobese African Americans: the National Heart, Lung, and Blood Institute Family Blood Pressure Program. *Am J Hypertens* **17**, 834-8 (2004).
25. Melton, L.J., 3rd. History of the Rochester Epidemiology Project. *Mayo Clin Proc* **71**, 266-74 (1996).
26. Patsch, J.R. *et al.* Relation of triglyceride metabolism and coronary artery disease. Studies in the postprandial state. *Arterioscler Thromb* **12**, 1336-45 (1992).
27. Lai, C.Q. *et al.* Fenofibrate effect on triglyceride and postprandial response of apolipoprotein A5 variants: the GOLDN study. *Arterioscler Thromb Vasc Biol* **27**, 1417-25 (2007).
28. Sorlie, P.D. *et al.* Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol* **20**, 629-41 (2010).
29. Lavange, L.M. *et al.* Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol* **20**, 642-9 (2010).
30. Laurie, C.C. *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet* **44**, 642-50 (2012).
31. Laurie, C.C. *et al.* Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol* **34**, 591-602 (2010).
32. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
33. Williams, R.R. *et al.* NHLBI family blood pressure program: methodology and recruitment in the HyperGEN network. Hypertension genetic epidemiology network. *Ann Epidemiol* **10**, 389-400 (2000).
34. Freedman, B.I. *et al.* Genome-wide scans for heritability of fasting serum insulin and glucose concentrations in hypertensive families. *Diabetologia* **48**, 661-8 (2005).
35. Pedersen, C.B., Gotzsche, H., Moller, J.O. & Mortensen, P.B. The Danish Civil Registration System. A cohort of eight million persons. *Dan Med Bull* **53**, 441-9 (2006).

36. Sebastiani, P. *et al.* A family longevity selection score: ranking sibships by their longevity, size, and availability for study. *Am J Epidemiol* **170**, 1555-62 (2009).
37. Newman, A.B. *et al.* Health and function of participants in the Long Life Family Study: A comparison with other cohorts. *Aging (Albany NY)* **3**, 63-76 (2011).
38. Scottish Council for Research in Education. *The intelligence of Scottish children: A national survey of an age-group*, (University of London Press, London, 1933).
39. Deary, I.J., Whiteman, M.C., Starr, J.M., Whalley, L.J. & Fox, H.C. The impact of childhood intelligence on later life: Following up the Scottish Mental Surveys of 1932 and 1947. *J. Pers. Soc. Psychol.* **86**, 130-147 (2004).
40. Deary, I.J., Whalley, L.J. & Starr, J.M. *A Lifetime of Intelligence, Follow-up Studies of the Scottish Mental Surveys of 1932 and 1947*, (American Psychological Association, Washington, DC, 2009).
41. Deary, I.J. *et al.* The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr.* **7**, 28 (2007).
42. Hsueh, W.C. *et al.* Diabetes in the Old Order Amish: characterization and heritability analysis of the Amish Family Diabetes Study. *Diabetes Care* **23**, 595-601 (2000).
43. Streeten, E.A. *et al.* Reduced incidence of hip fracture in the Old Order Amish. *J Bone Miner Res* **19**, 308-13 (2004).
44. Sorkin, J. *et al.* Exploring the genetics of longevity in the Old Order Amish. *Mech Ageing Dev* **126**, 347-50 (2005).
45. Mitchell, B.D. *et al.* The genetic response to short-term interventions affecting cardiovascular function: rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. *Am Heart J* **155**, 823-8 (2008).
46. Shuldiner, A.R. *et al.* Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA* **302**, 849-57 (2009).
47. Hofman, A. *et al.* The Rotterdam Study: 2016 objectives and design update. *European journal of epidemiology* **30**, 661-708 (2015).
48. Völzke et al. Cohort profile: the study of health in Pomerania. Int J Epidemiol. Apr;40(2):294-307 (2011).
49. Ridker, P.M. *et al.* Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. *Clin Chem* **54**, 249-55 (2008).
50. Grove, M.L. *et al.* Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* **8**, e68095 (2013).
51. Goldstein, J.I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543-5 (2012).
52. Barve, R.A. *et al.* Genetic association of left ventricular mass assessed by M-mode and two-dimensional echocardiography. *J Hypertens* **34**, 88-96 (2016).