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

Citation for published version:

Digital Object Identifier (DOI):
10.1016/j.ajhg.2018.12.001

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
American Journal of Human Genetics

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.
Associations of Mitochondrial and Nuclear Mitochondrial Variants and Genes with Seven Metabolic Traits


(introduction 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 CHARGE mtDNA+ 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 followed by single-variant and gene-based association testing. We report two significant common variants, one in MT-ATP6 and another in D-loop 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 ≤ 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 ≤ 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 membrane 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 inherited2–5 ~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 for protein translation. We study the genetic associations of mtDNA and nuclear genes (MT-nDNA) with seven metabolic traits, including BMI, WHR, glucose, insulin, HOMA-B, HOMA-IR, and HbA1c. This 45-cohort collaboration included 70,775 (insulin) to 170,202 (BMI) pan-ancestry individuals. We performed single-variant and gene-based association testing and assessed regulatory features of variants in the 109 genes associated with at least one of our seven traits. We also performed enrichment and protein-protein interactions analyses. Our results suggest that mtDNA and nuclear genes make important contributions to glucose and insulin metabolism, adipocyte regulation, diabetes, and cardiovascular disease.
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. 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.

The American Journal of Human Genetics 104, 112–138, January 3, 2019

(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.\(^1,12\) Because of heteroplasmacy, 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 196,976 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 (\(\mu\)U/mL),

---

medications. The participating cohorts were required to have at least
HbA1c values of glucose, insulin, and HbA1c to missing for T2D individuals
non-diabetic participants when studying glycemic traits, by setting
Methods, section 1). The association test analysis was restricted to
We observed differences in the raw measures of traits (Table S1).
All cohorts used the same units for all traits for the raw measures.
Preparation of Traits for Analyses
HOMA-B, HOMA-IR, and glycated hemoglobin (HbA1c, %) indica-
tive 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 insu-
lin, 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.
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 ≤
1E−03) are annotated with red arrows and gray text for single SNV associations
and with green arrow for gene-based associ-
cation, and text in red color represents
SNVs that passed the threshold p ≤
5E−04. The asterisk (*) marks missense or
synonymous variants; a dagger (†) marks
a SKAT gene-based significant association.
The mtDNA is shown as a single mole-
cule, 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-ND1, and then recy-
cling again from the first blue color at
MT-ND1. The full information is summa-
rized 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-RNKR1 [12S RNA] and MT-RNKR2
[16S RNA]), and 22 transport RNAs (MT-
TE, TV, TL1, T1, TQ, TM, TW, TA, TN, TC,
TY, TS1, TD, TK, TG, TR, TH, TS2, TL2,
TE, TT, TP).
mtDNA Variant Harmonization
All mtDNA variants from each array (Table S14) were annotated to
the nucleotide position according to the rCRS of the human mito-
chondrial 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 heterozy-
gotes 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 geno-
types 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
Age2 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.

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 pro-
gramming languages was used to produce residuals fitting the
following model:

\[
\text{trait} = \text{age} + \text{age}^2 + \text{sex} + \text{PCs} + \text{other study specific covariates } (\text{e.g., study center, etc})
\]
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). 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 sampled populations (the “Cosmopolitan” reference panel). In 1000G. The 1000G included 2,504 individuals representing sampled populations (the “Cosmopolitan” reference panel). 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 (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 ≥ 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 MAP^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) (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 (see Web Resources). Familial and maternal correlation structures 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, 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: \[ \text{FreqWeight}_{k} = \frac{\sum_i w_i f_i}{\sum_i w_i} \] where \( k \) represents the indexing of contributing studies, \( w_i \) is the sample size (number of individuals) per study, and \( f_i \) is the coded allele frequency for a particular marker from a particular study. Our working group conducted an internal permutation test 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 ≤ 0.05/49 ≤ 1E–03, a threshold for common variants (MAF ≥ 1%), but possibly liberal for rare variants (MAF < 1%) (Table 1). Furthermore, we considered p ≤ 0.05/(49 × 7 traits) ≤ 1.5E–04 as a conservative threshold when accounting for seven traits tested. We settled for a Bonferroni-threshold p ≤ 0.05/(49 × 2 domains) ≤ 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 ≥ 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 ≤ 1E–03) associated with 6 metabolic traits (Figure 1), by adding the filter of MAC ≥ 5 for each cohort, then 7 variants passed p ≤ 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. 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. The meta-analyses were performed using seqMeta (see Web Resources). The gene-based Bonferroni p value was calculated as p ≤ 0.05/ (37 mtDNA genes) ≤ 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, (2) Literature Lab (Acumen Biotech), and (3) MitoMiner (4.0) (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>
<thead>
<tr>
<th>No</th>
<th>Pos</th>
<th>Gene</th>
<th>rID</th>
<th>Annotation</th>
<th>Trait</th>
<th>A1/2</th>
<th>Freq1 WeightAve</th>
<th>FreqSE</th>
<th>MinFreq</th>
<th>MaxFreq</th>
<th>MAF</th>
<th>MAC</th>
<th>Dir</th>
<th>Het-p</th>
<th>N</th>
<th>Meta-analysis Level Results Selected with MAF of Weighted Allele Average Frequency &lt; 1% and Passing Meta-p-Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8076</td>
<td>MT-ATP6</td>
<td>rs1870651176</td>
<td>missense</td>
<td>WHR</td>
<td>A/G</td>
<td>0.9676</td>
<td>0.0295</td>
<td>0.9135</td>
<td>0.9974</td>
<td>0.0324</td>
<td>834</td>
<td></td>
<td>0.13 (0.04)</td>
<td></td>
<td>4.1E/C0</td>
</tr>
<tr>
<td>2</td>
<td>1830</td>
<td>MT-ND5</td>
<td>rs202120082</td>
<td>missense</td>
<td>INS</td>
<td>T/C</td>
<td>0.0158</td>
<td>0.0954</td>
<td>0.0028</td>
<td>0.2439</td>
<td>0.0158</td>
<td>301</td>
<td></td>
<td>--</td>
<td></td>
<td>--/C0</td>
</tr>
<tr>
<td>3</td>
<td>14124</td>
<td>MT-ND6</td>
<td>rs2853814</td>
<td>missense</td>
<td>BMI</td>
<td>T/C</td>
<td>0.0035</td>
<td>0.0029</td>
<td>0.0003</td>
<td>0.0022</td>
<td>0.0003</td>
<td>134</td>
<td></td>
<td>0.84 (0.24)</td>
<td></td>
<td>4.9E/C0</td>
</tr>
<tr>
<td>4</td>
<td>14353</td>
<td>MT-ND6</td>
<td>rs1559610</td>
<td>missense</td>
<td>BMI</td>
<td>T/C</td>
<td>0.0012</td>
<td>0.0007</td>
<td>0.0009</td>
<td>0.0022</td>
<td>0.0012</td>
<td>7</td>
<td></td>
<td>0.84 (0.24)</td>
<td></td>
<td>4.9E/C0</td>
</tr>
<tr>
<td>5</td>
<td>18694</td>
<td>MT-ND6</td>
<td>rs370513314</td>
<td>missense</td>
<td>BMI</td>
<td>T/C</td>
<td>0.0916</td>
<td>0.0032</td>
<td>0.0080</td>
<td>0.0978</td>
<td>0.0084</td>
<td>120</td>
<td></td>
<td>0.84 (0.24)</td>
<td></td>
<td>4.9E/C0</td>
</tr>
</tbody>
</table>

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^3 N; 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).

**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,\(^{31-50}\) 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 x 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 ≤ 0.05/(542 x 7 traits) ≤ 1.3E-05 could have been used, which is larger than the threshold we used, p ≤ 1E-06. Furthermore, to compare whether our MT-nDNA genes pass the genome-wide threshold of p ≤ 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
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 ≥ 0.80) with the best eSNV of the target gene. Depending on r^2 ≥ 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 ≤ 5E–04, see in Material and Methods: mtDNA Meta-Analyses). MT-8706 in MT-ATP6 (MIM: 516005) (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 ≤ 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.
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: 613221), IGFI (MIM: 147440), and BCL2 (MIM: 151430) were associated with more than one trait (Figures 2, 3, 8, and 9). 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” term (Table S7) while 18 were associated with “Cardiovascular disease” (Table S7). (For space limitation, a detailed enrichment analysis is provided in the Supplemental Material and Methods, section 8.) These
<table>
<thead>
<tr>
<th>No</th>
<th>SNV</th>
<th>Chrom</th>
<th>Position</th>
<th>Role</th>
<th>Gene</th>
<th>Full Gene Name (Cytogenetic Position)</th>
<th>BMI p</th>
<th>OBESITY p</th>
<th>WHR p</th>
<th>GLUC p</th>
<th>INS p</th>
<th>HOMA-B p</th>
<th>HOMAIR p</th>
<th>HbA1c p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs622798</td>
<td>1</td>
<td>45549599 upstream-2 kb</td>
<td>AKR1A1</td>
<td>Aldo-Keto Reductase Family 1 Member A1 (1p34.1)</td>
<td>8.56E-07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>rs1280316</td>
<td>1</td>
<td>66843725 Intron</td>
<td>WDR78</td>
<td>WD repeat domain 78 (1p31.3)</td>
<td>6.81E-07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>rs1093013</td>
<td>1</td>
<td>75634658 Intron</td>
<td>SLC44A5</td>
<td>Solute carrier family 44 member 5 (1p31.1)</td>
<td>-</td>
<td>-</td>
<td>4.86E-17</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>rs6428792</td>
<td>1</td>
<td>119114244 Intron</td>
<td>WARS2</td>
<td>Tryptophanyl tRNA Synthetase 2, Mitochondrial (1p12)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.95E-18</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>rs2301453</td>
<td>1</td>
<td>172389027 Intron</td>
<td>DNLM3</td>
<td>Dynamin 3 (1q24.3)</td>
<td>-</td>
<td>-</td>
<td>4.38E-17</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs4844390</td>
<td>1</td>
<td>207761504 Intron</td>
<td>CD46</td>
<td>CD46 Molecule, Complement Regulatory Protein (1q22.2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.90E-07</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>rs11118296</td>
<td>1</td>
<td>219408638 Intron</td>
<td>LYPLAL1</td>
<td>Lysophospholipase Like 1 (1q41)</td>
<td>-</td>
<td>-</td>
<td>3.00E-07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>rs6713865</td>
<td>2</td>
<td>23676937 Intron</td>
<td>KLHL29</td>
<td>Kelch like family member 29 (2p24.1)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.39E-13</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs934778</td>
<td>2</td>
<td>25166355 Intron</td>
<td>POMC</td>
<td>Proopiomelanocortin (2p23.3)</td>
<td>7.15E-07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs1107238</td>
<td>2</td>
<td>26235376 Intron</td>
<td>HADHA</td>
<td>Hydroxyacyl-CoA Dehydrogenase/3-Ketoacyl-CoA Thiolase/Enoyl-CoA Hydratase (Trifunctional Protein), Alpha Subunit (2p23.3)</td>
<td>8.86E-07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>rs13404446</td>
<td>2</td>
<td>27296386 intron</td>
<td>TRIMS4</td>
<td>Tripartite Motif Containing 54 (2p23.3)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.09E-09</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>rs4665965</td>
<td>2</td>
<td>27313513 intron</td>
<td>MPV17</td>
<td>Mitochondrial Inner Membrane Protein (2p23.3)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.83E-09</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>rs3736594</td>
<td>2</td>
<td>27772914 intron</td>
<td>MRPL33</td>
<td>Mitochondrial Ribosomal Protein L33 (2p23.2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.02E-13</td>
<td>4.67E-07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>rs4346434</td>
<td>2</td>
<td>43992607 intron</td>
<td>LRPPRC</td>
<td>Leucine Rich Pentatricopeptide Repeat Containing (2p21)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.52E-08</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs16843390</td>
<td>2</td>
<td>209655109 intron</td>
<td>MAP2</td>
<td>Microtubule Associated Protein 2 (2q24)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.32E-08</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs715</td>
<td>2</td>
<td>210678331 utr-3-prime</td>
<td>CPS1</td>
<td>Carbamoyl-Phosphate Synthase 1 (2q34)</td>
<td>5.81E-07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>rs933994</td>
<td>2</td>
<td>218785893 intron</td>
<td>CYP27A1</td>
<td>Cytochrome P450 Family 27 Subfamily A Member 1 (2q35)</td>
<td>8.65E-07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>rs17036328</td>
<td>3</td>
<td>12348985 intron</td>
<td>PPARG</td>
<td>Peroxisome Proliferator Activated Receptor Gamma (3p23.2)</td>
<td>4.27E-07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.59E-12</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>No</th>
<th>SNV</th>
<th>Chrom</th>
<th>Position</th>
<th>Role</th>
<th>Gene</th>
<th>Full Gene Name (Cytogenetic Position)</th>
<th>BMI p</th>
<th>OBESITY p</th>
<th>WHR p</th>
<th>GLUC p</th>
<th>INS p</th>
<th>HOMA-B p</th>
<th>HOMAIR p</th>
<th>HbA1c p</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>rs3729931</td>
<td>3</td>
<td>12585017</td>
<td>intron</td>
<td>RAF1</td>
<td>Raf-1 Proto-Oncogene, Serine/Threonine Kinase (3p25.2)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>20</td>
<td>rs11715915</td>
<td>3</td>
<td>49417897</td>
<td>nc-transcript</td>
<td>AMT</td>
<td>Aminomethyltransferase (3p21.31)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>21</td>
<td>rs12489828</td>
<td>3</td>
<td>52532998</td>
<td>intron</td>
<td>NT5DC2</td>
<td>5'-Nucleotide Domain Containing 2 (3p21.1)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>22</td>
<td>rs2365389</td>
<td>3</td>
<td>61250788</td>
<td>intron</td>
<td>FHIT</td>
<td>Fragile Histidine Triad (3p14.2)</td>
<td>3.75E–15</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>23</td>
<td>rs332375</td>
<td>3</td>
<td>66377079</td>
<td>intron</td>
<td>SLC25A26</td>
<td>Solute Carrier Family 25 Member 26 (3p14.1)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>24</td>
<td>rs1735536</td>
<td>3</td>
<td>128377770</td>
<td>intron</td>
<td>EFSEC</td>
<td>Eukaryotic Elongation Factor, Selenocysteine-TRNA Specific (3q21.3)</td>
<td>2.95E–10</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>25</td>
<td>rs9844666</td>
<td>3</td>
<td>136255374</td>
<td>intron</td>
<td>PCCB</td>
<td>Propionyl-CoA Carboxylase Beta Subunit (3q22.3)</td>
<td>7.22E–07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>26</td>
<td>rs11924648</td>
<td>3</td>
<td>171000207</td>
<td>intron</td>
<td>SLC2A2</td>
<td>Solute Carrier Family 2 Member 2 (3q26.2)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>27</td>
<td>rs10012946</td>
<td>4</td>
<td>6291623</td>
<td>intron</td>
<td>WFS1</td>
<td>Wolframin ER Transmembrane Glycoprotein (4p16.1)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>28</td>
<td>rs10518406</td>
<td>4</td>
<td>122841742</td>
<td>intron</td>
<td>FGF2</td>
<td>Fibroblast Growth Factor 2 (4q28.1)</td>
<td>6.50E–07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>29</td>
<td>rs1458758</td>
<td>4</td>
<td>122914724</td>
<td>intron</td>
<td>NUDT6</td>
<td>Nudix Hydrolase 6 (4q28.1)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>30</td>
<td>rs303084</td>
<td>4</td>
<td>123145793</td>
<td>intron</td>
<td>SPATA5</td>
<td>Spermatogenesis associated 5 (4q28.1)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>31</td>
<td>rs12654264</td>
<td>5</td>
<td>75352778</td>
<td>intron</td>
<td>HMGCR</td>
<td>3-Hydroxy-3-Methylglutaryl-CoA Reductase (5q13.3)</td>
<td>1.81E–08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>32</td>
<td>rs10478424</td>
<td>5</td>
<td>119453325</td>
<td>intron</td>
<td>HSD17B4</td>
<td>Hydroxysteroid 17-Beta Dehydrogenase 4 (5q23.1)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>33</td>
<td>rs2881156</td>
<td>5</td>
<td>135812973</td>
<td>intron</td>
<td>SLC25A48</td>
<td>Solute carrier family 25 member 48 (5q31.1)</td>
<td>–</td>
<td>8.10E–07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>34</td>
<td>rs3828870</td>
<td>6</td>
<td>16743066</td>
<td>intron</td>
<td>ATXN1</td>
<td>Ataxin 1 (6p22.3)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>35</td>
<td>rs1800562</td>
<td>6</td>
<td>26092913</td>
<td>intron</td>
<td>HFE</td>
<td>Hemochromatosis (6p22.2)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>36</td>
<td>rs1800626</td>
<td>6</td>
<td>31575254</td>
<td>upstream-2KB</td>
<td>TNF</td>
<td>Tumor necrosis factor (6p21.33)</td>
<td>–</td>
<td>7.30E–07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>37</td>
<td>rs6457796</td>
<td>6</td>
<td>34860776</td>
<td>intron</td>
<td>UHRF1BP1</td>
<td>UHRF1 binding protein 1 (6p21.31)</td>
<td>1.15E–09</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>No</td>
<td>SNV</td>
<td>Chrom</td>
<td>Position</td>
<td>Role</td>
<td>Gene</td>
<td>Full Gene Name (Cytogenetic Position)</td>
<td>BMI p</td>
<td>OBESITY p</td>
<td>WHR p</td>
<td>GLUC p</td>
<td>INS p</td>
<td>HOMA-B p</td>
<td>HOMAIR p</td>
<td>HbA1c p</td>
</tr>
<tr>
<td>----</td>
<td>-------</td>
<td>-------</td>
<td>----------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>--------------------------------------</td>
<td>-------</td>
<td>-----------</td>
<td>-------</td>
<td>--------</td>
<td>------</td>
<td>----------</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>38</td>
<td>rs10434</td>
<td>6</td>
<td>43785475 utr-3-prime</td>
<td>VEGFA</td>
<td>Vascular endothelial growth factor A (6p21.1)</td>
<td>8.80E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>rs1049354</td>
<td>6</td>
<td>88143732 utr-3-prime</td>
<td>CNR1</td>
<td>Cannabinoid receptor 1 (6q15)</td>
<td>9.57E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>rs9400239</td>
<td>6</td>
<td>108656460 intron</td>
<td>FOXO3</td>
<td>Forkhead box O3 (6q21)</td>
<td>1.61E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>rs1273733</td>
<td>6</td>
<td>121113149 downstream-500B</td>
<td>TBC1D32</td>
<td>TBC1 domain family member 32 (6q22.31)</td>
<td>3.96E−12</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>rs1049349</td>
<td>6</td>
<td>121449496 utr-3-prime</td>
<td>GJA1</td>
<td>Gap junction protein alpha 1; synonymous: CX43 (6q22.31)</td>
<td>4.18E−15</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>rs1293954</td>
<td>6</td>
<td>151669826 intron</td>
<td>ESR1</td>
<td>Estrogen receptor 1 (6q25.1-q25.2)</td>
<td>4.41E−09</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>rs1203576</td>
<td>7</td>
<td>40808233 intron</td>
<td>SUGCT C7orf10 (Succinyl-CoA:Glutarate-CoA Transferase, 7p14.1)</td>
<td>1.48E−10</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>rs2908289</td>
<td>7</td>
<td>44184343 intron</td>
<td>GCK</td>
<td>Glucokinase (7p13)</td>
<td>3.32E−08</td>
<td>6.00E−09</td>
<td>–</td>
<td>2.24E−19</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>rs1088667</td>
<td>7</td>
<td>44705214 intron</td>
<td>OGDH Oxoglutarate dehydrogenase (7p13)</td>
<td>7.90E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>rs16892421</td>
<td>8</td>
<td>106499705 intron</td>
<td>OXR1 Oxidation resistance 1 (8q23.1)</td>
<td>6.94E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>rs7835803</td>
<td>8</td>
<td>120030213 intron</td>
<td>DEPTOR</td>
<td>DEP domain containing MTOR interacting protein (8q24.12)</td>
<td>4.97E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>rs2777795</td>
<td>9</td>
<td>104910084 intron</td>
<td>ABCA1</td>
<td>ATP binding cassette subfamily A member 1 (9q31.1)</td>
<td>3.13E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>rs7023913</td>
<td>9</td>
<td>128255683 downstream-500B</td>
<td>DNMI</td>
<td>Dynamin 1 (9q34.11)</td>
<td>7.30E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>rs3829109</td>
<td>9</td>
<td>136362314 intron</td>
<td>DNLZ DNL-Type Zinc Finger (9q34.3)</td>
<td>1.13E−10</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>rs1244497</td>
<td>10</td>
<td>7838019 intron</td>
<td>TAF3</td>
<td>TATA-box binding protein associated factor 3 (10p14)</td>
<td>1.84E−11</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>rs5030913</td>
<td>10</td>
<td>69246375 intron</td>
<td>HKDC1</td>
<td>Hexokinase domain containing 1 (10q22.1)</td>
<td>3.56E−13</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>rs4745982</td>
<td>10</td>
<td>69330087 intron</td>
<td>HK1</td>
<td>Hexokinase 1 (10q22.1)</td>
<td>2.87E−65</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>rs7899106</td>
<td>10</td>
<td>85651147 intron</td>
<td>GRID1</td>
<td>Glutamate Ionotropic Receptor Delta Type Subunit 1 (10q23.2)</td>
<td>2.96E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>rs7917772</td>
<td>10</td>
<td>102727686 intron</td>
<td>SFXN2</td>
<td>Sideroflexin 2 (10q24.32)</td>
<td>1.45E−09</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>rs1004467</td>
<td>10</td>
<td>102834750 intron</td>
<td>CYP17A1</td>
<td>Cytochrome P450 Family 17 Subfamily A Member 1 (10q24.32)</td>
<td>1.18E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>SNV</td>
<td>Chrom</td>
<td>Position</td>
<td>Role</td>
<td>Gene</td>
<td>Full Gene Name (Cytogenetic Position)</td>
<td>BMI p</td>
<td>OBESITY p</td>
<td>WHR p</td>
<td>GLUC p</td>
<td>INS p</td>
<td>HOMA-B p</td>
<td>HOMAIR p</td>
<td>HbA1c p</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>-------</td>
<td>----------</td>
<td>-------------------</td>
<td>------------------------------------</td>
<td>---------------------------------------</td>
<td>--------</td>
<td>-----------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
<td>----------</td>
<td>----------</td>
<td>---------</td>
</tr>
<tr>
<td>58</td>
<td>rs3740390</td>
<td>10</td>
<td>102878723</td>
<td>intron</td>
<td>AS3MT</td>
<td>Arsenite Methyltransferase (10q24.32)</td>
<td>4.82E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>59</td>
<td>rs4758633</td>
<td>11</td>
<td>219538</td>
<td>intron</td>
<td>SIRT3</td>
<td>Sirtuin 3 (11p15.5)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>3.44E−10</td>
</tr>
<tr>
<td>60</td>
<td>rs757110</td>
<td>11</td>
<td>17396930</td>
<td>missense</td>
<td>ABCC8</td>
<td>ATP Binding Cassette Subfamily C Member 8 (11p15.1)</td>
<td>4.23E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>61</td>
<td>rs10767664</td>
<td>11</td>
<td>27704439</td>
<td>intron</td>
<td>BDNF</td>
<td>Brain Derived Neurotrophic Factor (11p14.1)</td>
<td>5.53E−13</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>62</td>
<td>rs11038913</td>
<td>11</td>
<td>46538180</td>
<td>intron</td>
<td>AMBRA1</td>
<td>Autophagy and beclin 1 regulator (11p11.2)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>63</td>
<td>rs11039149</td>
<td>11</td>
<td>47255124</td>
<td>intron</td>
<td>NR1H3</td>
<td>Nuclear Receptor Subfamily 1 Group H Member 3 (11p11.2)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>64</td>
<td>rs7118178</td>
<td>11</td>
<td>47637583</td>
<td>intron</td>
<td>MTCH2</td>
<td>Mitochondrial carrier 2 (11p11.2)</td>
<td>5.12E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>65</td>
<td>rs4246215</td>
<td>11</td>
<td>61796827</td>
<td>utr-3-prime</td>
<td>FEN1</td>
<td>Flap Structure-Specific Endonuclease 1 (11p12.2)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>66</td>
<td>rs174556</td>
<td>11</td>
<td>61813163</td>
<td>intron</td>
<td>FADS1</td>
<td>Fatty acid desaturase 1 (11q12.2)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>67</td>
<td>rs7943191</td>
<td>11</td>
<td>62561079</td>
<td>intron</td>
<td>EF1G</td>
<td>Eukaryotic translation elongation factor 1 gamma (11q12.3)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>68</td>
<td>rs11231150</td>
<td>11</td>
<td>62584330</td>
<td>intron</td>
<td>TUT1</td>
<td>Terminal uridylyl transferase 1, U6 snRNA-specific (11q12.3)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>69</td>
<td>rs1017639</td>
<td>11</td>
<td>68831066</td>
<td>intron</td>
<td>CPT1A</td>
<td>Carnitine Palmitoyltransferase 1A (11q13.3)</td>
<td>4.96E−10</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>70</td>
<td>rs1296252</td>
<td>11</td>
<td>83273268</td>
<td>intron</td>
<td>CCDC90B</td>
<td>Coiled-Coil Domain Containing 90B (11q13.3)</td>
<td>7.60E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>71</td>
<td>rs2110073</td>
<td>12</td>
<td>6966719</td>
<td>intron</td>
<td>PHB2</td>
<td>Prohibitin 2 (12p13.31)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>4.44E−08</td>
</tr>
<tr>
<td>72</td>
<td>rs7311050</td>
<td>12</td>
<td>7013532</td>
<td>intron</td>
<td>LPCAT3</td>
<td>Lysophosphatidylcholine Acyltransferase 3 (12p13.31)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>8.60E−08</td>
</tr>
<tr>
<td>73</td>
<td>rs1049380</td>
<td>12</td>
<td>26336611</td>
<td>downstream-500B</td>
<td>ITPR2</td>
<td>Inositol 1,4,5-Trisphosphate Receptor Type 2 (12p11.23)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>74</td>
<td>rs2408955</td>
<td>12</td>
<td>48105348</td>
<td>upstream-2KB</td>
<td>PFKM</td>
<td>Phosphofructokinase, muscle (12q13.11)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.42E−15</td>
</tr>
<tr>
<td>75</td>
<td>rs35767</td>
<td>12</td>
<td>102481791</td>
<td>missense</td>
<td>IGF1</td>
<td>Insulin Like Growth Factor 1 (12q23.2)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>7.27E−08</td>
<td>7.57E−08</td>
<td>–</td>
</tr>
<tr>
<td>76</td>
<td>rs4766578</td>
<td>12</td>
<td>111466567</td>
<td>intron</td>
<td>ATXN2</td>
<td>Ataxin 2 (12q24.12)</td>
<td>2.85E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.84E−07</td>
</tr>
<tr>
<td>77</td>
<td>rs9581856</td>
<td>13</td>
<td>27451478</td>
<td>upstream-2KB</td>
<td>MTIF3</td>
<td>Mitochondrial Translational Initiation Factor 3 (13q12.2)</td>
<td>1.03E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>No</th>
<th>SNV</th>
<th>Chrom</th>
<th>Position</th>
<th>Role</th>
<th>Gene</th>
<th>Full Gene Name (Cytogenetic Position)</th>
<th>BMI p</th>
<th>OBESITY p</th>
<th>WHR p</th>
<th>GLUC p</th>
<th>INS p</th>
<th>HOMA-B p</th>
<th>HOMAIR p</th>
<th>HbA1c p</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>rs1124607</td>
<td>13</td>
<td>27921083</td>
<td>intron</td>
<td>PDX1</td>
<td>Pancreatic And Duodenal Homeobox 1 (13q12.2)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>4.49E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>79</td>
<td>rs1325363</td>
<td>13</td>
<td>33192439</td>
<td>intron</td>
<td>STARD13</td>
<td>STAR related lipid transfer domain containing 13 (13q13.1-q13.2)</td>
<td>7.25E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>80</td>
<td>rs1078892</td>
<td>13</td>
<td>40563883</td>
<td>intron</td>
<td>FOXO1</td>
<td>Forkhead Box O1 (13q14.11)</td>
<td>5.11E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>81</td>
<td>rs7143963</td>
<td>14</td>
<td>102838088</td>
<td>intron</td>
<td>TRAF3</td>
<td>TNF Receptor Associated Factor 3</td>
<td>2.82E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>82</td>
<td>rs12908437</td>
<td>15</td>
<td>98744146</td>
<td>intron</td>
<td>IGF1R</td>
<td>Insulin Like Growth Factor 1 Receptor</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>6.32E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>83</td>
<td>rs740862</td>
<td>16</td>
<td>3639677</td>
<td>intron</td>
<td>DNASE1</td>
<td>Deoxyribonuclease 1</td>
<td>3.21E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>84</td>
<td>rs151181</td>
<td>16</td>
<td>28479196</td>
<td>intron</td>
<td>CLN3</td>
<td>CLN3, Battenin</td>
<td>2.10E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>85</td>
<td>rs8055138</td>
<td>16</td>
<td>28880144</td>
<td>intron</td>
<td>ATP2A1</td>
<td>ATPase Sarcoplasmic/ Endoplasmic Reticulum Ca2+ Transporting 1</td>
<td>8.17E−17</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>86</td>
<td>rs749767</td>
<td>16</td>
<td>31113086</td>
<td>downstream-500B</td>
<td>BCKDK</td>
<td>Branched Chain Ketoacid Dehydrogenase Kinase</td>
<td>1.21E−09</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>87</td>
<td>rs7186084</td>
<td>16</td>
<td>68782357</td>
<td>intron</td>
<td>CDH1</td>
<td>Cadherin 1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.09E−07</td>
<td>–</td>
</tr>
<tr>
<td>88</td>
<td>rs1847591</td>
<td>16</td>
<td>78908913</td>
<td>intron</td>
<td>WWOX</td>
<td>WW domain containing oxidoreductase</td>
<td>–</td>
<td>–</td>
<td>9.99E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>89</td>
<td>rs9904685</td>
<td>17</td>
<td>1352101</td>
<td>intron</td>
<td>YWHAE</td>
<td>Tyrosine 3-monooxygenase/ tryptophan 5-monooxygenase activation protein epsilon</td>
<td>6.27E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>90</td>
<td>rs4646404</td>
<td>17</td>
<td>17516885</td>
<td>intron</td>
<td>PEMT</td>
<td>Phosphatidylethanolamine N-Methyltransferase</td>
<td>–</td>
<td>–</td>
<td>5.30E−11</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>91</td>
<td>rs9914988</td>
<td>17</td>
<td>28856086</td>
<td>intron</td>
<td>ERAL1</td>
<td>Era like 125 mitochondrial tRNA chaperone 1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>92</td>
<td>rs242559</td>
<td>17</td>
<td>45948522</td>
<td>intron</td>
<td>MAPT</td>
<td>Microtubule Associated Protein Tau</td>
<td>–</td>
<td>–</td>
<td>8.29E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>93</td>
<td>rs1319247</td>
<td>17</td>
<td>63106279</td>
<td>intron</td>
<td>TANC2</td>
<td>Etratricopeptide repeat, ankyrin repeat and coiled-coil containing 2</td>
<td>–</td>
<td>–</td>
<td>6.02E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>94</td>
<td>rs12940622</td>
<td>17</td>
<td>80641771</td>
<td>intron</td>
<td>RPTOR</td>
<td>Regulatory Associated Protein Of MTOR Complex 1</td>
<td>2.49E−09</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>95</td>
<td>rs1044661</td>
<td>17</td>
<td>82943144</td>
<td>intron</td>
<td>B3GNTL1/TBCD</td>
<td>UDP-GlcNAc:BetaGal Beta-1,3-N-Acetylglicosaminyltransferase Like 1/Tubulin Folding CofactorD</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.74E−46</td>
</tr>
<tr>
<td>96</td>
<td>rs1788785</td>
<td>18</td>
<td>23562376</td>
<td>intron</td>
<td>NPC1</td>
<td>NPC Intracellular Cholesterol Transporter 1</td>
<td>1.98E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>No</th>
<th>SNV</th>
<th>Chrom</th>
<th>Position</th>
<th>Role</th>
<th>Gene Name (Cytogenetic Position)</th>
<th>BMI p</th>
<th>OBESITY p</th>
<th>WHR p</th>
<th>GLUC p</th>
<th>INS p</th>
<th>HOMA-B p</th>
<th>HOMAIR p</th>
<th>HbA1c p</th>
</tr>
</thead>
<tbody>
<tr>
<td>97</td>
<td>rs17066842</td>
<td>18</td>
<td>60373391</td>
<td>upstream-2KB</td>
<td>MC4R Melanocortin 4 Receptor</td>
<td>6.40E−14</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>98</td>
<td>rs12454712</td>
<td>18</td>
<td>63178651</td>
<td>intron</td>
<td>BCL2 BCL2, Apoptosis Regulator</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.10E−09</td>
<td>1.39E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>99</td>
<td>rs757318</td>
<td>18</td>
<td>18709498</td>
<td>intron</td>
<td>CRTC1 CREB Regulated Transcription Coactivator 1</td>
<td>8.76E−09</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>100</td>
<td>rs2075650</td>
<td>19</td>
<td>44892362</td>
<td>intron</td>
<td>TOMM40 Translocase Of Outer Mitochondrial Membrane 40</td>
<td>1.25E−08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>101</td>
<td>rs405509</td>
<td>19</td>
<td>44905579</td>
<td>upstream-2 kb</td>
<td>APOE Apolipoprotein E</td>
<td>2.65E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>102</td>
<td>rs2281361</td>
<td>20</td>
<td>32140338</td>
<td>intron</td>
<td>TM9SF4 ransmembrane 9 superfamily member 4</td>
<td>9.78E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>103</td>
<td>rs878639</td>
<td>20</td>
<td>35306660</td>
<td>intron</td>
<td>UQCC1 Ubiquinol-Cytochrome C Reductase Complex Assembly Factor 1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.50E−11</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>104</td>
<td>rs2076574</td>
<td>20</td>
<td>41092733</td>
<td>intron</td>
<td>TOP1 DNA topoisomerase I</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>4.11E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>105</td>
<td>rs5750373</td>
<td>22</td>
<td>37028990</td>
<td>intron</td>
<td>MPST Mercaptopynuate Sulfurtransferase</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2.17E−07</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>106</td>
<td>rs2284099</td>
<td>22</td>
<td>43155850</td>
<td>intron</td>
<td>TSPO Translocator Protein</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>6.65E−07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>107</td>
<td>rs1050828</td>
<td>23</td>
<td>154536002</td>
<td>missense</td>
<td>G6PD Glucose-6-phosphate dehydrogenase</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>8.23E−135</td>
</tr>
<tr>
<td>108</td>
<td>rs1448032</td>
<td>23</td>
<td>155052530</td>
<td>intron</td>
<td>FUNDC2 FUN14 domain containing 2</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>5.92E−17</td>
</tr>
<tr>
<td>109</td>
<td>rs5940514</td>
<td>23</td>
<td>155559972</td>
<td>intron</td>
<td>TMLHE Trimethyllysine hydroxylase, epsilon</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2.22E−19</td>
</tr>
</tbody>
</table>

p values in italics annotate SNVs that associate as significant with more than one trait.
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. 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 GWAS p value for the MT-nDNA SNVs with additional evidence from trans-gene expression regulation for a specific variant using GWAS3D83 (Figures S12.1–S12.3). The GWAS3D software selected 16 cell types, which included chromosomal looping data (SC 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 receptors86 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.
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 < 0.05), 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 < 1E−06 but did not reach p < 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. These rare mitochondrial genetic diseases commonly result in multiple clinical phenotypes with varying penetrance, likely due to differences in heteroplasmy. 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. 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, 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)) 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

Figure 3. Glucose and Insulin Association Results with MT-nDNA Candidate Genes
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 ≤ 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. 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 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 rs290828 (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, r² = 0.65, p = 8E–56),55 with T2D (rs1799884, r² = 0.81, p = 5E–18),103 and with metabolic syndrome (rs3757840, r² = 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 MVP17 (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 phosphates 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.51 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])14 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 MEC2P (Xq28 [MIM: 300005]) (Figure S12.3).117 MEC2P binds to methylated DNA and mediates transcriptional repression through interaction with histone deacetylase and the co-repressor 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 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 TEAM (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; WHR: rs10465617, TFB2M, p = 3.1E–03; glucose: rs4804124, DNMT1, p = 7.3E–05; insulin: rs7253062, DNMT1, p = 4.4E–03; HOMA-B: rs12462004, DNMT1, p = 4.4E–03; HOMA-IR: rs892189, DNMT1, p = 7E–04; HbA1c: rs11006132, TFAM, p = 2.4E–03).

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 mitochondrial tRNA synthase. We observed an intronic rs6428792 of WARS2 in association with WHR (p = 7.95E–18). Another SNV, rs10923724 in LD with rs6428792, r2 = 0.29, WHR p = 9E–25) 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.

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. NPC1 (a cholesterol transporter) is highly expressed in human white adipose tissue adipocytes with increased levels in obese individuals. 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. 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. 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, adrenocorticotropin 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., Locke et al., and Winkler et al. 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. and 16th in Locke et al. 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. The MTCH2 GWAS-variant rs7118178 (Table 4) has been associated with BMI, glucose, and pro-insulin (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. 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 CIQTNF4 (11p11.2 [MIM: 614911]) and PSFN3 (11p11.2 [MIM: 186852]) in subcutaneous adipose tissue. rs3817335, which is in high LD (r2 = 0.903) with rs4752856, regulates SLC39A13 (11p11.2 [MIM: 607385]) in tibial artery tissue, while rs7118178 (in less LD to two other SNVs, r2 = 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). CIQTNF4 is reported as a potential cytokine...
that can induce the activation of both NFKB1 (MIM: 164011) and IL6/STAT3 (MIM: 147620/102582) signaling pathways and acts in the hypothalamus to modulate food intake and peripheral energy expenditure. 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.

CELF1 has been associated with Alzheimer disease and obesity. Consequently, MTCH2 SNVs regulate gene 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 r² 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).

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 for plotting regional information for MTCH2 selected SNVs, which have been reported by large GWAS publications (for BMI T1, for glucose T2, for proinsulin T3, and another large analysis for glucose 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 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 r² between the MTCH2 eSNV, for example, 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).

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. 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).

Recently, Francis et al. 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 power138 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 that the MT-candidates (SNVs/genes) have special putative regulatory genes expression. We documented in this study how 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

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.
Web Resources

1000 Genomes data (released April 2016), http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/
EasyQC, http://www.geneepi-regensburg.de/easyqc/
FCGENE, https://sourceforge.net/projects/fcgene/
GTEx (Gene Expression Portal), https://gtexportal.org
GWAS Catalog, http://www.ebi.ac.uk/gwas/
Human Mitochondrion Sequence, https://www.mitomap.org/
IMPUTE2, https://mathgen.stats.ox.ac.uk/impute/impute_v2.html
MAGIC consortium archives, https://www.magicinvestigators.org/downloads
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


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—NDR; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VAR; 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.


