

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

Jaakko Laaksonen*¹, Ilkka Seppälä¹, Emma Raitoharju¹, Nina Mononen¹, Leo-Pekka Lyytikäinen¹, Melanie Waldenberger^{2,3}, Thomas Illig^{4,5}, Maija Lepistö⁶, Henriikki Almusa⁶, Pekka Ellonen⁶, Nina Hutri-Kähönen⁷, Markus Juonala^{8,9,10}, Mika Kähönen¹¹, Olli Raitakari^{12,13}, Jukka T. Salonen^{14,15}, Terho Lehtimäki¹

¹Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

²Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany

³Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany

⁴Hannover Unified Biobank, Hannover Medical School, Hannover, Germany

⁵Institute for Human Genetics, Hannover Medical School, Hannover, Germany

⁶Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland

⁷Department of Pediatrics, Tampere University Hospital and Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

⁸Department of Medicine, University of Turku, Turku, Finland

⁹Division of Medicine, Turku University Hospital, Turku, Finland

¹⁰Murdoch Children's Research Institute, Parkville, Victoria, Australia

¹¹Department of Clinical Physiology, Tampere University Hospital and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

¹²Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku
University Hospital, Turku, Finland

¹³Research Centre for Applied and Preventive Cardiovascular Medicine, University of Turku,
Turku, Finland

¹⁴Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland

¹⁵MAS-Metabolic Analytical Services Oy, Helsinki, Finland

*Corresponding author

Jaakko Laaksonen

Faculty of Medicine and Health Technology

PO Box 100

FI-33014 Tampere University, Finland

Tel.: +358 504080774

Fax: +358 32599003

e-mail: jaakko.h.laaksonen@tuni.fi

Abstract

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

Methods: The genome-wide blood cell expression data of 19,637 probes, 199 deep-sequenced mtSNPs, and nine haplogroups of 955 individuals from a population-based Young Finns Study cohort were used. Significant associations were identified with linear regression and analysis of covariance. The effects of sex and prediabetes on the associations between gene expression and mtSNPs were studied using random-effect meta-analysis.

Results: Our analysis identified 53 significant expression probe-mtSNP associations after Bonferroni correction, involving 7 genes and 31 mtSNPs. Eight probe-mtSNP signals remained independent after conditional analysis. In addition, five genes showed differential expression between haplogroups. The meta-analysis did not show any significant differences in linear model effect sizes between males and females but identified the association between the *OASL* gene and mtSNP C16294T to show prediabetes-specific effects.

Conclusions: This study pinpoints new independent mtSNPs associated with peripheral blood transcriptomics and replicates six previously reported associations, providing further evidence of the mitochondrial genetic control of blood cell gene expression. In addition, we present evidence that prediabetes might lead to perturbations in mitochondrial control.

Introduction

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

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

Interaction among mtDNA and nuclear DNA encoded factors is important to cellular function and in adaptation to environmental changes. Since the co-evolution of these two genomes has been demonstrated at a protein-protein interaction level and some modes of co-regulation at a

transcriptional level also exist, it is logical that disruption in mitochondria-nuclear signaling may lead to disease. (8, 9) Some mtSNPs have been reported to be associated with type 2 diabetes (T2D), but this association has been seen only in Asians (10–12). A Korean study found mitochondrial haplogroups to be associated with an increased or decreased risk of T2D and with altered nuclear gene expression patterns that correlate with the susceptibility to develop T2D (13). However, it is unclear to what extent the mitochondrial-nuclear interaction is altered when the glucose homeostasis has already been impaired. Therefore, in addition to sex-specific differences, of special interest in this study is prediabetes, a precursor state and a major risk factor for the development of T2D (14).

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

Results

Effect of mtSNPs and haplogroups on peripheral blood gene expression

A total of 3,907,763 expression probe-mtSNP pairs were tested for association in a linear regression model. The genomic inflation factor was 1.00 (quantile-quantile plot for expected versus observed p-values shown in Supplementary material, Figure S1), indicating that an inflation of the genetic association due to population stratification or undetected systematic error is unlikely. As shown in Table 1, a total of 53 expression probe-mtSNP pairs were significant after Bonferroni correction, corresponding to five nuclear and two mitochondrial genes and 31 mtSNPs. These seven identified genes regulated by mtSNPs were signal peptidase complex subunit 2 pseudogene 4 (*SPCS2P4*), ring finger protein 113A (*RNF113A*), signal peptidase complex subunit 2 (*SPCS2*), mitochondrially

encoded cytochrome c oxidase II (*MT-CO2*), cardiolipin synthase 1 (*CRLS1*), solute carrier family 25 member 15 (*SLC25A14*), and mitochondrially encoded 16S RNA like 1 (*MT-RNR2L1*).

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

ANCOVA and Tukey's post hoc test indicated that five expression probes, corresponding to four nuclear and one mitochondrial genes, showed differential expression between haplogroups ($\lambda_{GC} = 1.04$, Supplementary Figure S2). Three of the genes, *SPCS2P4*, *RNF113A*, and *SLC25A14*, were also associated with individual mtSNPs, but the other two genes, solute carrier family 2 member 8 (*SLC2A8*, Illumina Array Address 5870326) and mitochondrially encoded NADH dehydrogenase 5 (*MT-ND5*, Illumina Array Address 4880477), were not identified in the probe-mtSNP analysis. The results of Tukey's post hoc test are shown in Table 3. Figure 2 illustrates the expression levels of the top two genes, *SPCS2P4* and *RNF113A*, which were significantly lower in haplogroup K compared to all other eight major haplogroups. The expression levels of the other three genes across haplogroups are shown in Supplementary Figure S3.

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

Sex- and prediabetes-specific effects

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

Discussion

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

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

After applying pairwise conditional analysis to all mtSNPs showing at least modest linkage disequilibrium, five probe-mtSNP pairs showed evidence for independent signaling. In the expression of *SPCS2P4*, the effects of G9055A and A11251G/C15452T were in opposite directions which also supports the independence of these signals. For *SLC25A14* and *CRLS1*, none of the signals remained independent. This could be explained by the high linkage disequilibrium (pairwise $r^2 > 0.89$ for all mtSNPs), all variants that associated with these two genes in the unconditioned analysis together predispose to differential gene expression. In addition, the associations with mitochondrial genes *MT-CO2* and *MT-RNR2L1* seem to result from independent signals. That is to say, in total, eight mtSNP-gene associations were independent.

We also discovered that the expression intensities of four nuclear and one mitochondrial genes differed between the major Scandinavian haplogroups present in our study population. The haplogroup distribution shown in Table 5 is similar to the one previously reported in Finnish population. Compared to other European populations, the frequency of haplogroup U appears to be higher and the frequency of haplogroup K lower in Finland. (3, 15) The lower intensities of *SPCS2P4* and *RNF113A* in haplogroup K result from the fact that the majority of the mtSNPs associated with these two genes (e.g. G9055A, A3480G, and A10550G) are also the defining variants for this haplogroup (16). The expression intensities of two genes, *SLC2A8* and *MT-ND5*, were not associated at the mtSNP level. However, the majority of the mtSNPs (C295T, C462T, A12612G, G13708A, C16069T) defining haplogroup J (16) that associated with *MT-ND5*, were in the top 20 variants

associating with this gene. Those associations were not however significant after correction for multiple testing. For haplogroup U that associated with *SLC2A8*, none of the defining variants were in top 20 mtSNPs associating with this gene. This implies that the haplogroup effect is driven by small individual effects of many mtSNPs, especially the expression of *SLC2A8*.

As already discussed earlier (7), the biological relevance of these reported associations remain unclear, and they do not necessarily imply causal relationships—i.e. all these mtSNPs are not necessarily expression regulatory SNPs. However, the associations may equally represent the altered cellular activities resulting from the mtSNPs that the nuclear genome is then compensating for (13). In other studies, some of the reported mtSNPs have been associated with non-mitochondrial diseases. For example, the variant G9055A and haplogroup K have been found to increase breast cancer risk in European-American women (17). Interestingly, increased levels of protein coded by *RNF113A* in plasma have been suggested to act as a biomarker for breast cancer (18), while our analysis showed a reduced expression of *RNF113A* associated with this mtSNP and haplogroup.

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

Finally, we tested the hypothesis that probe-mtSNP associations are affected by the onset of prediabetes. We were able to find support for this hypothesis by comparing the effect sizes between samples with and without prediabetes, which showed that the onset of prediabetes affects the

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

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

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

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

Materials and Methods

Study participants

The Cardiovascular Risk in Young Finns Study (YFS, <http://youngfinnsstudy.utu.fi>) is a Finnish longitudinal population study on the evolution of cardiovascular risk factors from childhood to adulthood (27). We used data from the follow-up in 2011 when the subjects were aged between 34 and 49 years, with the exception of the mtDNA data, which was gathered in 2007. The follow-up studies in 2007 and 2011 included 2,204 and 2,060 participants, respectively. The study plan was approved by the ethics committees of all participating hospital districts, and the study protocol of

each study phase corresponded with the proposal by the World Health Organization. All subjects gave written informed consent, and the study was conducted in accordance with the Declaration of Helsinki.

Blood transcriptomic analysis, RNA analysis, and data processing

Whole-blood samples were obtained from 2,049 individuals for RNA isolation. 63 samples were discarded during the RNA isolation protocol, leading to 1,987 samples including one technical replicate taking part in the genome wide expression analysis. 322 samples had too low concentration after amplification step. After this, 1,667 samples (including three replicates, two from the mRNA amplification step) were analyzed with an Illumina HumanHT-12 version 4 Expression BeadChip (Illumina Inc.) containing 47,231 expression and 770 control probes. The transcripts detected (detection p-value < 0.01) in less than 5 % of samples were excluded from the analysis. After this filtering 19,637 genes were used for analysis. We disregarded four samples with less than 6,000 significantly detected expression probes (detection p-value < 0.01).

The expression data was processed in R (<http://www.r-project.org/>) using a nonparametric background correction, followed by quantile normalization with control and expression probes, using the `neqc` function in the `limma` package (28) and a \log_2 transformation. The expression levels were also zero-centered, and rank-based inverse normal transformation was applied to further normalize the expression levels. Based on RPS4Y1–2 and XIST mRNA levels on the Y and X chromosomes, respectively, we excluded nine samples due to mismatch with the recorded sex. After quality control, expression data were available for 1,654 samples including four technical replicates, which were used to examine batch effects and excluded subsequently. In summary, the expression analysis was successful for 19,637 probes and 1,650 samples. Other details of the process have been described previously by Turpeinen et al. (29).

MtDNA sequencing

Genomic DNA sample (n = 1,817) concentrations were measured from whole-blood samples with the Qubit BR dsDNA kit (Life Technologies Ltd). Mitochondrial DNA was amplified from the genomic DNA using the REPLI-g Mitochondrial DNA kit (Qiagen) in a 15 µl reaction volume. After the enrichment, the amplified mtDNA samples were processed into Illumina deep sequencing compatible libraries with the Nextera DNA sample preparation kit (Illumina Inc.). The mtDNA concentrations were measured with Qubit dsDNA for Nextera tagmentation reaction. The reaction volume in the Nextera tagmentation and amplification steps was 20 µl, and after both steps the libraries were purified with a EdgeBio Performa V3 96-Well Short Plate (Edge BioSystems). After the amplification, the libraries were first incubated with 4 µl of EdgeBio SOPE Resin and then purified with EdgeBio Performa plates. After purification, 48 samples with different index tags were pooled together (2 µl each) in each pool and concentrated with DNA Clean & ConcentratorTM-5 (Zymo Research). The final volume of the concentrated pool was 15 µl. The sequencing-ready libraries were quantitated with an Agilent 2100 Bioanalyzer High Sensitivity kit (Agilent). The libraries were deep-sequenced with the Illumina HiSeq system. All the samples (n = 1,667) that achieved any mean bait coverage were included in the quality control process.

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

MtDNA quality control and data processing

First, samples with individual missingness > 0.10 (134 samples) were excluded. Next, samples were classified into haplogroups by using HaploGrep (32) (Phylotree build 16) (16) after comparison to the revised Cambridge Reference Sequence (rCRS) (1). Only those individuals whose haplogroup quality score was above 0.90 were included for further analyses (66 samples rejected). At this quality threshold, haplogroup assignment is quite reliable, according to the HaploGrep manual. After this, 1,467 samples remained for further analysis. Then, samples that had both gene expression and mtDNA data were merged, after which 955 samples were remaining. Next, 29 mtSNPs that obtained mean call rates of < 0.85 were excluded, all samples had mean call rate above this value. After this, heteroplasmic alleles' status were set to missing. The remaining quality control steps included filtering for missingness by mtSNP > 0.05 (34 mtSNPs discarded), and minor allele frequency (MAF) < 0.01 (538 mtSNPs discarded). After these procedures, a total of 199 mtSNPs from 955 samples (548 women and 407 men) were available for probe-mtSNP association analysis (Supplementary Material, Table S2).

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

Definition of prediabetes

Venous blood samples were drawn after an overnight fast for the determination of serum glucose and glycated hemoglobin A1c (HbA1c). The classification of prediabetes was based on fasting glucose and HbA1c according to the criteria of the American Diabetes Association (14). People with impaired fasting glucose were defined as having a fasting plasma glucose level of 5.6–6.9 mmol/L or an HbA1c level of 39–47 mmol/mol without a diagnosis of T2D. The diagnosis of T2D included subjects with a fasting plasma glucose level of 7.0 mmol/L or higher or an HbA1c level of 48 mmol/mol or higher,

or those who reported using oral glucose-lowering medication or insulin (but had not reported having type 1 diabetes), or who had a reported diagnosis of T2D by a physician. Those diagnosed with type 1 diabetes were also ruled out. Of the 833 subjects for whom gene expression and mtDNA data and prediabetes status were available, 249 had prediabetes, and 584 controls had normal levels of fasting plasma glucose and HbA1c.

Statistical analysis

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

The sex-specific effects of mtSNPs on gene expression were tested by applying the same linear model as described above (sex was removed from the covariates) to males and females separately. Differences in effect sizes were compared by applying a random-effect meta-analytic model to each probe using the MetaDE package (34). Heterogeneity was examined by Cochran's Q test with the corresponding p-value. A significant p-value would suggest that there is a significant difference in effect sizes between the sexes. For the sex-specific meta-analysis, the number of mtSNPs tested was

156 because for some mtSNPs the minor allele frequency was less than 0.01 in either males or females (Supplementary Material, Table S3). Significance was then defined as $p < 1.63 \times 10^{-8}$ (i.e. $0.05/[156 \times 19,637]$).

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

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

Both nuclear and mitochondrial PCs were added as covariates, in addition to those mentioned above, in the linear mtSNP association models until no additional reduction in genomic inflation

factor (λ_{GC}) could be achieved. Nuclear PCs 1–11 and mitochondrial PCs 1–2 were used for all probe-mtSNP association analyses. For haplogroup analysis in ANCOVA, nuclear PCs 1–11 were used. Mitochondrial PCs cannot be applied to haplogroup analysis, since the haplogroups are strongly correlated with the mitochondrial PCs. The genomic inflation factor was calculated using the GenABEL package (38). Values of $\lambda_{GC} < 1.05$ are generally considered benign (39).

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

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

Data availability

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

Acknowledgements

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

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

Conflict of interest statement

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

References

1. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. and Howell, N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.*, **23**, 147.
2. Kivisild, T., Shen, P., Wall, D.P., Do, B., Sung, R., Davis, K.K., Passarino, G., Underhill, P.A., Scharfe, C., Torroni, A., *et al.* (2005) The role of selection in the evolution of human mitochondrial genomes. *Genetics*, **172**, 373–387.
3. Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.-L. and Wallace, D.C. (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics*, **144**, 1835–1850.
4. Hurst, L.D. and Frank, S.A. (1996) Mitochondria and male disease. *Nature*, **383**, 224.
5. Gemmell, N.J., Metcalf, V.J. and Allendorf, F.W. (2004) Mother's curse: the effect of mtDNA on individual fitness and population viability. *Trends Ecol. Evol.*, **19**, 238–244.
6. Innocenti, P., Morrow, E.H. and Dowling, D.K. (2011) Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. *Science*, **332**, 845–848.
7. Kassam, I., Qi, T., Lloyd-Jones, L., Holloway, A., Bonder, M.J., Henders, A.K., Martin, N.G., Powell, J.E., Franke, L., Montgomery, G.W., *et al.* (2016) Evidence for mitochondrial genetic control of autosomal gene expression. *Hum. Mol. Genet.*, **25**, 5332–5338.
8. Woodson, J.D. and Chory, J. (2008) Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.*, **9**, 383–395.
9. Bar-Yaacov, D., Blumberg, A. and Mishmar, D. (2012) Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation. *Biochim. Biophys. Acta*, **1819**, 1107–1111.

10. Chinnery, P.F., Elliott, H.R., Patel, S., Lambert, C., Keers, S.M., Durham, S.E., McCarthy, M.I., Hitman, G.A., Hattersley, A.T. and Walker, M. (2005) Role of the mitochondrial DNA 16184-16193 poly-C tract in type 2 diabetes. *Lancet*, **366**, 1650–1651.
11. Park, K.S., Chan, J.C., Chuang, L.M., Suzuki, S., Araki, E., Nanjo, K., Ji, L., Ng, M., Nishi, M., Furuta, H., *et al.* (2008) A mitochondrial DNA variant at position 16189 is associated with type 2 diabetes mellitus in Asians. *Diabetologia*, **51**, 602–608.
12. Saxena, R., De Bakker, P.I.W., Singer, K., Mootha, V., Burt, N., Hirschhorn, J.N., Gaudet, D., Isomaa, B., Daly, M.J., Groop, L., *et al.* (2006) Comprehensive Association Testing of Common Mitochondrial DNA Variation in Metabolic Disease. *Am. J. Hum. Genet.*, **79**, 54–61.
13. Hwang, S., Kwak, S.H., Bhak, J., Kang, H.S., Lee, Y.R., Koo, B.K., Park, K.S., Lee, H.K., Cho, Y.M., Heon Kwak, S., *et al.* (2011) Gene expression pattern in transmitochondrial cytoplasmic hybrid cells harboring type 2 diabetes-associated mitochondrial DNA haplogroups. *PLoS One*, **6**, e22116.
14. American Diabetes Association (2010) Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, **33**, S62-S69.
15. Richards, M., Côrte-Real, H.C., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H.-J. and Sykes, B. (1996) Paleolithic and Neolithic Lineages in the European Mitochondrial Gene Pool. *Am. J. Hum. Genet.*, **59**, 185–203.
16. Van Oven, M. and Kayser, M. (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.*, **30**, E386–E394.
17. Bai, R.-K., Leal, S.M., Covarrubias, D., Liu, A. and Wong, L.-J.C. (2007) Mitochondrial genetic background modifies breast cancer risk. *Cancer Res.*, **67**, 4687–4694.

18. Li, C.I., Mirus, J.E., Zhang, Y., Ramirez, A.B., Ladd, J.J., Prentice, R.L., McIntosh, M.W., Hanash, S.M. and Lampe, P.D. (2012) Discovery and preliminary confirmation of novel early detection biomarkers for triple-negative breast cancer using preclinical plasma samples from the Women's Health Initiative observational study. *Breast Cancer Res. Treat.*, **135**, 611–618.
19. Ebner, S., Mangge, H., Langhof, H., Halle, M., Siegrist, M., Aigner, E., Paulmichl, K., Paulweber, B., Datz, C., Sperl, W., *et al.* (2015) Mitochondrial haplogroup T is associated with obesity in Austrian juveniles and adults. *PLoS One*, **10**, e0135622.
20. Mueller, E.E., Eder, W., Ebner, S., Schwaiger, E., Santic, D., Kreindl, T., Stanger, O., Paulweber, B., Iglseider, B., Oberkofler, H., *et al.* (2011) The Mitochondrial T16189C Polymorphism Is Associated with Coronary Artery Disease in Middle European Populations. *PLoS One*, **6**, e16455.
21. Ghosh, A.R., Bhattacharya, R., Bhattacharya, S., Nargis, T., Rahaman, O., Duttagupta, P., Raychaudhuri, D., Shiu, C., Liu, C., Roy, S., *et al.* (2016) Adipose Recruitment and Activation of Plasmacytoid Dendritic Cells Fuel Metaflammation. *Diabetes*, **65**, 3440–3452.
22. Ferrannini, E., Gastaldelli, A. and Iozzo, P. (2011) Pathophysiology of Prediabetes. *Med. Clin. North Am.*, **95**, 327–339.
23. Festa, A., Hanley, A.J.G., Tracy, R.P., D'agostino, R. and Haffner, S.M. (2003) Inflammation in the Prediabetic State Is Related to Increased Insulin Resistance Rather Than Decreased Insulin Secretion. *Circulation*, **108**, 1822–1830.
24. Peltonen, M., Korpi-Hyövälti, E., Oksa, H., Puolijoki, H., Saltevo, J., Vanhala, M., Saaristo, T., Saarikoski, L., Sundvall, J. and Tuomilehto, J. (2006) Prevalence of obesity, type 2 diabetes, and other disturbances in glucose metabolism in Finland - FIN-D2D survey. *Finnish Med. J.*, **61**, 163–170.

25. Centers for Disease Control and Prevention (2011) National Diabetes Fact Sheet: National Estimates and General Information on Diabetes and Prediabetes in the United States, 2011. *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, 2011.*
26. Tabák, A.G., Herder, C., Rathmann, W., Brunner, E.J. and Kivimäki, M. (2012) Prediabetes: a high-risk state for diabetes development. *Lancet*, **379**, 2279–2290.
27. Juonala, M., Viikari, J.S. and Raitakari, O.T. (2013) Main findings from the prospective Cardiovascular Risk in Young Finns Study. *Curr. Opin. Lipidol.*, **24**, 57–64.
28. Smyth, G.K. (2005) Limma: linear models for microarray data. In *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer, New York, NY, pp. 397–420.
29. Turpeinen, H., Seppälä, I., Lyytikäinen, L.L.-P., Raitoharju, E., Hutri-Kähönen, N., Levula, M., Oksala, N., Waldenberger, M., Klopp, N., Illig, T., *et al.* (2015) A genome-wide expression quantitative trait loci analysis of proprotein convertase subtilisin/kexin enzymes identifies a novel regulatory gene variant for FURIN expression and blood pressure. *Hum. Genet.*, **134**, 627–636.
30. Pietiläinen, K.H., Naukkarinen, J., Rissanen, A., Saharinen, J., Ellonen, P., Kernen, H., Suomalainen, A., Götz, A., Suortti, T., Yki-Järvinen, H., *et al.* (2008) Global transcript profiles of fat in monozygotic twins discordant for BMI: pathways behind acquired obesity. *PLoS Med.*, **5**, e51.
31. Sulonen, A.-M., Ellonen, P., Almusa, H., Lepistö, M., Eldfors, S., Hannula, S., Miettinen, T., Tyynismaa, H., Salo, P., Heckman, C., *et al.* (2011) Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol.*, **12**, R94.

32. Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G. and Kronenberg, F. (2011) HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.*, **32**, 25–32.
33. Dong, X., Chang, T.-W., Weiss, S.T. and Qiu, W. (2017) powerEQTL: Power and Sample Size Calculation for eQTL Analysis. R package version 0.1.3.
34. Wang, X., Kang, D.D., Shen, K., Song, C., Lu, S., Chang, L.-C., Liao, S.G., Huo, Z., Tang, S. and Ding, Y. (2012) An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. *Bioinformatics*, **28**, 2534–2536.
35. Schurmann, C., Heim, K., Schillert, A., Blankenberg, S., Carstensen, M., Dörr, M., Felix, S., Endlich, K., Gieger, C., Grallert, H., *et al.* (2012) Analyzing Illumina Gene Expression Microarray Data from Different Tissues: Methodological Aspects of Data Analysis in the MetaXpress Consortium. *PLoS One*, **7**, e50938.
36. Biffi, A., Anderson, C.D., Nalls, M.A., Rahman, R., Sonni, A., Cortellini, L., Rost, N.S., Matarin, M., Hernandez, D.G., Plourde, A., *et al.* (2010) Principal-component analysis for assessment of population stratification in mitochondrial medical genetics. *Am. J. Hum. Genet.*, **86**, 904–917.
37. Landgraf, A.J. and Lee, Y. (2015) Dimensionality reduction for binary data through the projection of natural parameters. *arXiv:1510.06112*.
38. Aulchenko, Y.S., Ripke, S., Isaacs, A. and Van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, **23**, 1294–1296.
39. Price, A.L., Zaitlen, N.A., Reich, D. and Patterson, N. (2010) New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.*, **11**, 459–463.

40. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

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

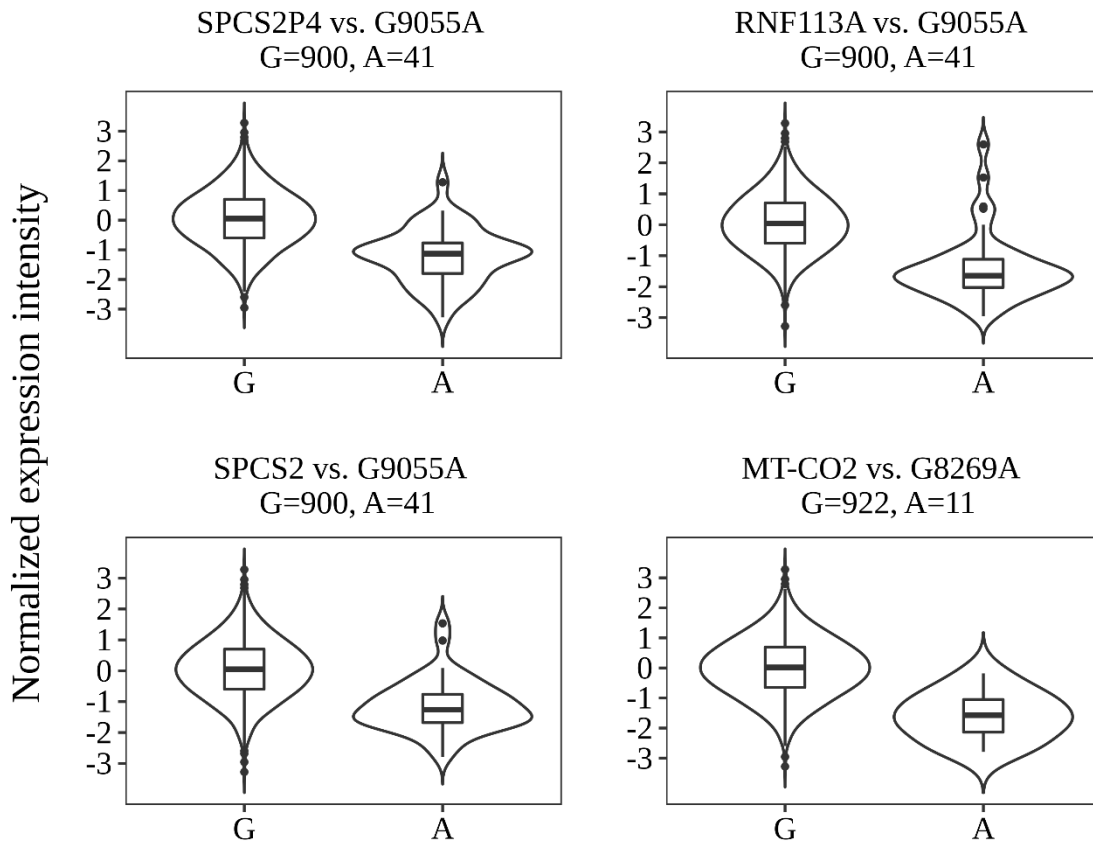


Figure 2. Combined boxplot and violin plot of the normalized expression intensities of *SPCS2P4* and *RNF113A* across the haplogroups.

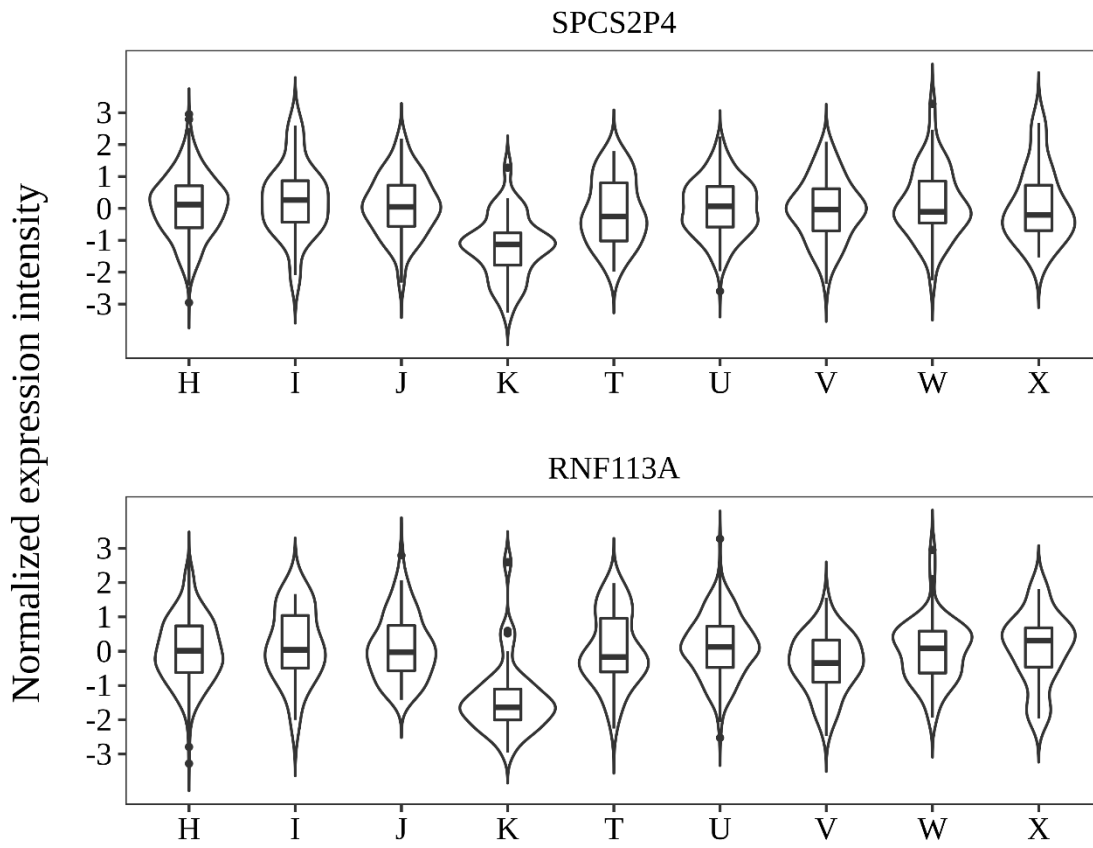


Figure 3. Combined boxplot and violin plot of the normalized expression intensities of *OASL* across the individuals with and without prediabetes relative to mtSNP C16294T.

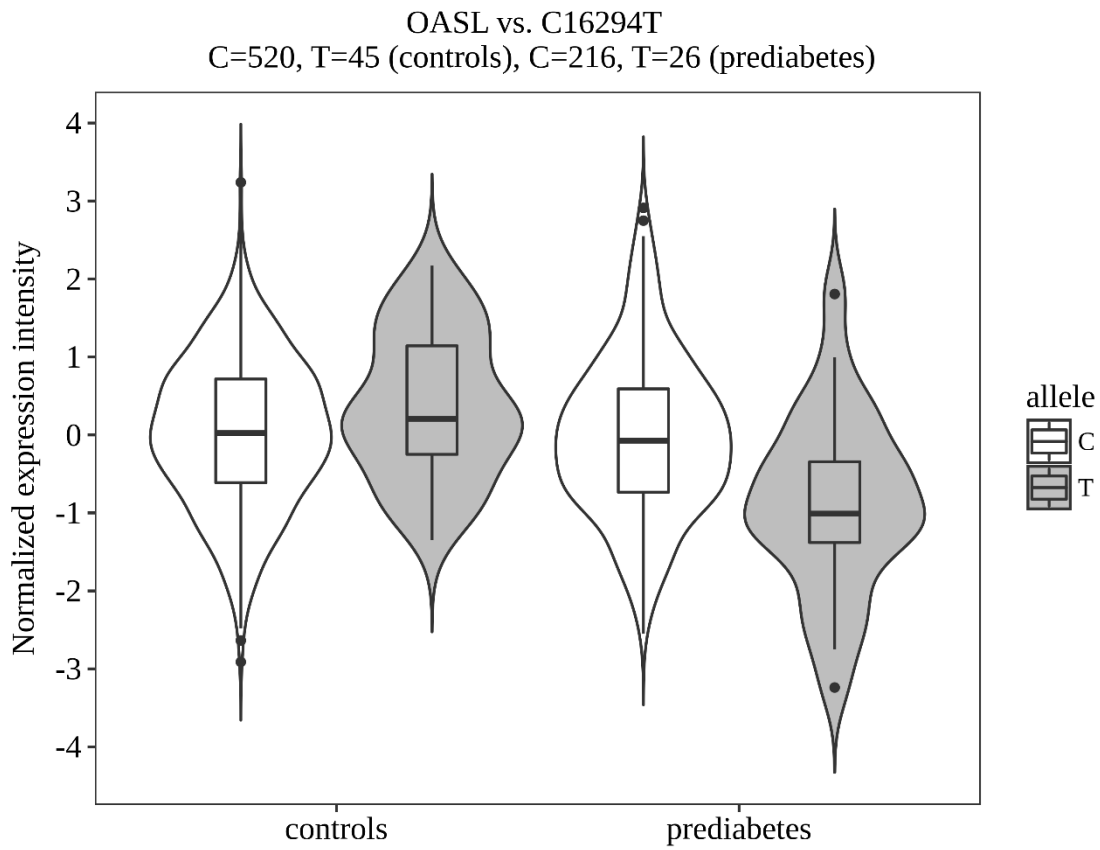


Table 1. 53 probe-mtSNP associations that had a p of $< 1.28 \times 10^{-8}$. The beta-coefficient represents the proportion of 1 SD change in normalized gene expression intensity (mean = 0, SD = 1). An asterisk (*) marks the associations that were replicated from the previous results (7).

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

| | | | | | | |
|----------|---------|---------|------|-------|------|-------------------------|
| SLC25A14 | 1710754 | G5046A | 0.05 | 0.74 | 0.12 | 1.76 x 10 ⁻⁹ |
| SPCS2P4 | 4210315 | A10398G | 0.14 | -0.40 | 0.07 | 1.80 x 10 ⁻⁹ |
| SLC25A14 | 1710754 | G15884C | 0.05 | 0.74 | 0.12 | 2.25 x 10 ⁻⁹ |
| RNF113A | 6940196 | A1811G | 0.08 | -0.60 | 0.10 | 3.43 x 10 ⁻⁹ |
| RNF113A | 6940196 | C15452A | 0.12 | 1.39 | 0.24 | 5.22 x 10 ⁻⁹ |
| MT-CO2 | 6550386 | A16162G | 0.06 | -0.66 | 0.11 | 5.48 x 10 ⁻⁹ |
| SPCS2 | 7040068 | T11299C | 0.04 | -0.79 | 0.13 | 7.40 x 10 ⁻⁹ |
| RNF113A | 6940196 | A11251G | 0.12 | 1.36 | 0.23 | 7.84 x 10 ⁻⁹ |
| SLC25A14 | 1710754 | T12414C | 0.05 | 0.69 | 0.12 | 8.01 x 10 ⁻⁹ |
| SPCS2 | 7040068 | T1189C | 0.03 | -0.82 | 0.14 | 8.27 x 10 ⁻⁹ |
| SLC25A14 | 1710754 | G5460A | 0.05 | 0.65 | 0.11 | 9.93 x 10 ⁻⁹ |
| RNF113A | 6940196 | G11377A | 0.03 | -0.93 | 0.16 | 1.18 x 10 ⁻⁸ |

Abbreviations: MAF, minor allele frequency; SE, standard error; SD, standard deviance

Table 2. Independent mtSNP-gene association signals from the pairwise conditional analysis. The beta-coefficient represents the proportion of 1 SD change in normalized gene expression intensity (mean = 0, SD = 1). The mtSNP-specific Bonferroni-corrected p-value accounts for the number of pairwise analyses made for each mtSNP and is defined as the limit of significance.

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

Abbreviations: SE, standard error; SD, standard deviance

Table 3. Haplogroup-wise comparisons from Tukey's post-hoc test for the five differentially expressed genes identified in ANCOVA. One unit of difference in means represents the proportion of 1 SD change in normalized gene expression intensity.

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

Abbreviations: CI, confidence interval; SD, standard deviance

Table 4. Characteristics of the population used in the prediabetes-specific analysis. Values are means (SD). The Mann-Whitney U test was used for age and body mass index and the χ^2 test for sex when the two groups were compared. P-value ≤ 0.001 for all comparisons.

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

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

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

Abbreviations

mtDNA, mitochondrial DNA; mtSNP, mitochondrial single nucleotide polymorphism; T2D, type 2 diabetes; YFS, Young Finns Study; MAF, minor allele frequency; HbA1c, glycated hemoglobin; ANCOVA, analysis of covariance; PC principal component; SE, standard error; SD, standard deviance; CI, confidence interval