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

Jaakko Laaksonen1, Ilkka Seppälä1, Emma Raitoharju1, Nina Mononen1, Leo-Pekka Lyytikäinen1, Melanie Waldenberger2,3, Thomas Illig2,4,5, Maija Lepistö6, Henrikki Alamusa6, Janne Asikainen7, Pekka Ellonen6, Nina Hutri-Kähönen8, Markus Juonala9,10, Mika Kähönen11, Olli Raitakari12,13, Jukka T. Salonen14,15, Terho Lehtimäki1

1Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Finland
2Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Germany
3Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Germany
4Hannover Unified Biobank, Hannover Medical School, Germany
5Institute for Human Genetics, Hannover Medical School, Germany
6Institute for Molecular Medicine (FIMM), University of Helsinki, Finland
7Department of Health, National Institute for Health and Welfare, Finland
8Department of Pediatrics, Tampere University Hospital and Faculty of Medicine and Life Sciences, University of Tampere, Finland
9Department of Medicine, University of Turku, Finland
10Division of Medicine, Turku University Hospital, Finland
11Department of Clinical Physiology, Tampere University Hospital and Finnish Cardiovascular Research Center – Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Finland
12Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Finland
13Research Centre for Applied and Preventive Cardiovascular Medicine, University of Turku, Finland
14Hjelt Institute, University of Helsinki, Finland
15MAS-Metabolic Analytical Services Oy, Finland

**Corresponding author:**Jaakko Laaksonen, BM, MSc
Faculty of Medicine and Life Sciences, Arvo building, room D339
PO Box 100
FI-33014 University of Tampere, Finland
Tel.: +358 504080774
e-mail: jaakko.laaksonen@uta.fi

**Abstract word count:** 244 **Word count:** 3,136

**Abstract**

**Background:** The effect of mitochondrial DNA variation affecting 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, the evidence is insufficient. 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 special respect to sex- and prediabetes-specific effects.

**Methods:** 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 showed 52 significant expression probe-mtSNP associations after Bonferroni correction, involving 7 genes and 31 mtSNPs. In addition, five genes showed differential expression between haplogroups. Meta-analysis did not show any significant differences in linear model effect sizes between males and females but identified the association between *OASL* gene and mtSNP C16294T to show prediabetes-specific effects.

**Conclusions:** This study pinpoints 46 new mtSNPs associated with peripheral blood transcriptomics and replicates six previously reported associations, providing further evidence for the mitochondrial genetic control of blood cell gene expression. In addition, evidence that prediabetes might lead to perturbations in this genetic control was observed.

**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 during evolution, these polymorphisms have accumulated dividing the human population into smaller mitochondrial haplogroups. The deﬁnition of these haplogroups is based on particular combinations of certain single-nucleotide polymorphisms (SNPs), which are scattered throughout the mitochondrial genome. 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 in *Drosophila melanogaster*, whichfound 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 in humans showed 15 significant associations between mtSNPs and nuclear gene expression but found only little evidence for 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 another mtSNPs not included in the study sample. In another gene expression study a single mtSNP and the depletion of mtDNA also resulted in specific changes in nuclear gene expression (8).

Since both nucleus and mitochondrion have proteins encoded by their genomes, it is physiologically essential that these two coordinate and coevolve in order to maintain suitable cellular function and adaptation to environmental changes. In cells, mitochondria-related signaling to the nucleus affects cellular activities under both normal and pathophysiological conditions. (9) Some mtSNPs have been reported to be associated with type 2 diabetes mellitus (T2DM) but this association has been seen only in Asians (10–12). A Korean study found out that mitochondrial haplogroups associated with increased or decreased risk of T2DM had altered nuclear gene expression patterns correlating to the susceptibility to T2DM (13). However, it is unclear to what extent the mitochondria-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 T2DM (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 investigate whether any associations show sex-specific differences, d) to study the differential gene expression between the major Scandinavian haplogroups present in our study population and finally e) to see whether any gene expression-mtSNP associations are affected by the onset of prediabetes.

**Materials and Methods**

*Study participants*

Gene expression and mtDNA data were from the Cardiovascular Risk in Young Finns Study (YFS, http://youngfinnsstudy.utu.fi) which is a Finnish longitudinal population study on the evolution of cardiovascular risk factors from childhood to adulthood (15). The present study included 955 individuals (548 women and 407 men) aged between 34 and 49 for whom complete gene expression and mtDNA data were available.

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*

RNA was isolated from whole-blood samples and the expression levels were analysed with an Illumina HumanHT-12 version 4 Expression BeadChip (Illumina Inc.) containing 47,231 expression and 770 control probes. 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 (16) and a log2 transformation. The expression levels were also zero-centered and rank-based inverse normal transformation was applied to further normalize the expression levels. The expression analysis was successful for 19,637 probes. Other details of the process have been previously described by Turpeinen et al. (17).

*MtDNA sequencing*

Genomic DNA sample concentrations were measured from whole-blood samples with Qubit BR dsDNA kit (Life Technologies Ltd). Mitochondrial DNA was amplified from the genomic DNA using REPLI-g Mitochondrial DNA kit (Qiagen) in 15µl reaction volume. After the enrichment the amplified mtDNA samples were processed to Illumina deep sequencing compatible libraries with Nextera DNA sample preparation kit (Illumina). The mtDNA concentrations were measured with Qubit dsDNA for Nextera tagmentation reaction. The reaction volume in Nextera tagmentation and amplification steps was 20µl and after both steps the libraries were purified with EdgeBio Performa V3 96-Well Short Plate (Edge BioSystems). After the amplification the libraries were first incubated with 4µl 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 15µl. The sequencing ready libraries were quantitated with Agilent 2100 Bioanalyzer High Sensitivity kit (Agilent). The libraries were deep sequenced in Illumina HiSeq system.

The primer sites in REPLI-g kit have not been published and therefore the data for each 16 samples was validated amplifying the mtDNA in two different PCR amplicons covering the whole mtDNA. The primers have been previously described by Pietiläinen et al. (18). The amplicons were processed to Illumina compatible sequencing libraries according to the same Nextera protocol as above. The data was analyzed using an in-house developed bioinformatics pipeline (19). The variants from REPLI-g amplification and PCR amplification were compared and no significant differences were observed. Comparison results confirm that REPLI-g primers do not affect the variant detection.

*MtDNA quality control and data processing*

First, samples and mtSNPs that obtained mean call rates < 0.85 and mtSNPs that included heteroplasmy were excluded. Remaining quality control steps included filtering for missingness by individual > 0.10, missingness by mtSNP > 0.05 and minor allele frequency (MAF) < 0.01. After these procedures, a total of 199 mtSNPs from 955 samples were available for mtSNP-probe association analysis (Supplementary Material, Table S1).

Classification into haplogroups was carried out by using HaploGrep (20) (Phylotree build 16) (21) after comparison to the revised Cambridge Reference Sequence (rCRS) (1). Only those individuals whose haplogroup quality score was above 0.90 were included into statistical analyses. At this quality threshold haplogroup assignment is quite reliable according to HaploGrep's manual. For association testing, haplogroups were assigned to major haplogroups. Haplogroups with frequency less than 0.01 were excluded, leaving 934 samples for the haplogroup-probe analysis. The haplogroup frequencies are shown in Table 1. All other major Scandinavian haplogroups are present except haplogroup M (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 to 6.9 mmol/L or HbA1c of 39 to 47 mmol/mol and not diagnosed with T2DM. The diagnosis of T2DM included subjects with a fasting plasma glucose level of 7.0 mmol/L or higher or HbA1c of 48 mmol/mol or higher or those with reported use of oral glucose-lowering medication or insulin (but had not reported having T1DM) or who had a reported diagnosis of T2DM by a physician. Also those diagnosed with T1DM were 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 fasting plasma glucose and Hba1c levels.

Statistical analysis

In order to investigate the association of peripheral blood gene expression with mtSNPs, the expression levels were modelled 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. 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 × 10-8 (i.e. 0.05/(199 × 19,637)). Analysis of covariance (ANCOVA) was employed to flag genes for those showing differential expression between haplogroups. All genes with a p-value < 2.55 × 10-6 (i.e. 0.05/19,637) were compared using Tukey’s honest significant difference test to confirm between-haplogroup differences. A Tukey-adjusted p-value < 0.05 was considered statistically significant.

Sex-specific effects of mtSNPs on gene expression were tested by applying the same linear model as above 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 (22). 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 between effect sizes between sexes. For 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 S2). Significance was then defined as p < 1.63 × 10-8 (i.e. 0.05/(156 × 19,637)).

The effect of prediabetes on the association between mtSNPs and gene expression was studied similarly. Age, sex and body mass index were added as covariates in 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 × 10-8 (i.e. 0.05/(127 × 19,637) (Supplementary Material, Table S3).

To account for possible population stratification or undetected systematic procedural error, principal component (PC) analysis was performed on all mtSNPs and nuclear probes passing quality control. The use of mitochondrial PCs effectively removes false-positive associations but does not remove those that are true-positive (23). The prcomp function (package stats) was used to calculate nuclear PCs 1-20 and the logisticPCA package (24) was used to extract the same number of mitochondrial PCs. Both nuclear and mitochondrial PCs were added as covariates in the linear regression/ANCOVA models until no additional reduction in genomic inflation factor (λGC) could be achieved. Nuclear PCs 1-11 and mitochondrial PCs 1-2 were used for all probe-mtSNP association analyses. For ANCOVA, nuclear PCs 1-11 and the first mitochondrial PC were used. Genomic inflation factorwas calculated using the GenABEL package (25). Values of λGC < 1.05 are generally considered benign (26).

**Results**

*Effect of mtSNPs and haplogroups to peripheral blood gene expression*

A total of 3,907,763 expression probe-mtSNP pairs were tested for association in a linear regression model. Genomic inflation factor was 0.99 (quantile-quantile plot for expected versus observed p-values shown in Supplementary material, Figure S1), indicating that the inflation of genetic association due to population stratification or undetected systematic error was unlikely. As shown in Table 2, a total of 52 expression probe-mtSNP pairs were significant after Bonferroni correction, corresponding for 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*). Figure 1 illustrates the normalized expression intensities for the top four genes relative to the alleles of the top associated mtSNP.

ANCOVA and Tukey’s post hoc test indicated that five expression probes, corresponding for four nuclear and one mitochondrial genes, showed differential expression between haplogroups (λGC = 1.03, 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 5870326) were not identified in the mtSNP-probe analysis. The results from Tukey’s post hoc test are shown in Table 3. Figure 2 illustrates the expression levels of the top two genes, *SPCS2P4* and *RNF113A*, whichwere significantly lower in haplogroup K compared to all other eight major haplogroups.

*Sex- and prediabetes-specific effects*

Random-effect meta-analysis showed no statistically significant differences in the effect sizes of gene expression between sexes (results not shown). For one mtSNP-probe pair the meta-analysis showed a significant difference in the effect sizes between subjects with prediabetes and controls. A p-value of 8.91 x 10-9 (λGC = 0.99, Supplementary Figure S3) 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 x 10-6 (λGC = 0.99, Supplementary Figure S4) while the control group had an effect estimate of 0.43 (standard error of 0.12) and a p-value of 4.22 x 10-4  (λGC = 1.00, Supplementary Figure S5). That is, on average, subjects with prediabetes and minor allele T had lower expression levels of *OASL* compared to the reference allele while subjects with T allele but no prediabetes had higher expression levels compared to the reference allele, as can be seen in Figure 3.

**Discussion**

Our analysis of genome-wide peripheral blood gene expression and mtDNA variation provided additional evidence for mitochondrial genetic control of gene expression in humans. From 3,907,763 analyzed expression probe-mtSNPs pairs, 52 significant associations were identified 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, the latter encodes a protein containing two zinc finger domains. As compared to the previous results (7), we were able to replicate six of 15 reported associations. In other words, we showed 46 new associations. The nucleotide sequence in these previous results (7) differs from our sequence by one position, mtSNP A3481G in previous sequence corresponds to A3480G in our sequence and A10551G corresponds to A10550G, respectively. In addition to mtSNPs, gene expression intensities differed also between haplogroups. 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 (21).

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 aren’t necessarily expression regulatory SNPs. However, these associations may as well represent the altered cellular activities resulting from the mtSNPs that the nuclear genome is then compensating (13). In other studies some of the reported mtSNPs have been associated with non-mitochondrial diseases. For example, variant G9055A and haplogroup K have been found to increase breast cancer risk in European-American women (27). Interestingly, increased levels of protein coded by *RNF113A* in plasma has been suggested to act as a biomarker for breast cancer (28) while our analysis showed reduced expression of *RNF113A* associated with this mtSNP and haplogroup.

Random-effect meta-analysis did not reveal any sex-specific differences in the mitochondrial genetic control of gene expression. However, it is possible that sexually dimorphic effects on gene expression are mediated via another mtSNPs that were not included in this study. It is also worth mentioning that, to our knowledge, the two studies that have examined these sex-specific effects in humans have only taken into account homoplasmic mtDNA alleles. For this reason, additional studies taking into account also heteroplasmic alleles are needed.

A comparison of effect sizes between samples with and without prediabetes showed some evidence that the onset of prediabetes affects the mitochondrial control of the expression of *OASL* through mtSNP C16294T. This mtSNP in mtDNA control region has been also previously associated with cardiovascular risk factors, it has been linked to obesity in Austrian population (29) and there is also evidence that C16294T is associated with coronary artery disease, although possibly through linkage to mtSNP T16189C (30). Functionally, *OASL* encodes an interferon-inducible antiviral protein and its high expression levels in visceral adipose tissue, together with other three interferon signature genes, has been found to be positively correlated with adipose tissue and systemic insulin resistance (31). These previous results link *OASL* to prediabetes, since most persons with prediabetes are also insulin-resistant (32). Our results are in opposite direction to those published earlier in adipose tissue (31) 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, like adipose tissue. Nevertheless, since the mechanisms that regulate gene expression in liver and mononuclear leukocytes are similar (33) and leukocytes can migrate from circulation to various tissues (34), the gene expression in leukocytes might reflect the expression also in adipose tissue. One hypothesis from our results would be that the chronic low-grade inflammation in prediabetes (35) is, to some extent, milder in individuals with mtSNP C16294T. Whether this has any clinical significance remains to be examined in longitudinal studies.

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 has also 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 for sex-specific effects of mtSNPs on gene expression were found but some evidence was observed that the onset of prediabetes may lead to perturbations in this genetic control. The possible clinical relevance of these results remains to be examined in future functional and longitudinal studies.

**Acknowledgements**

We thank Jari Kaikkonen for technical assistance.

**Funding**

This article was funded 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); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; Finnish Cultural Foundation – The Pirkanmaa Regional Fund; and EU Horizon 2020 (grant 755320 for TAXINOMISIS).

**Disclosure statement**

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

**Contribution statement**

JL contributed to the study design, statistical analyses, data interpretation and wrote the manuscript. IS contributed to the study design and statistical analyses. LPL, HA and JA contributed to the primary data analysis. ER, NM, MW, TI, ML, NHK and MJ contributed to the data collection. PE developed the sequencing method and contributed to primary data analysis. MK and JTS contributed to obtaining funding and cohort collection. OR leads YFS and contributed to obtaining funding and cohort collection. TL supervised the research and contributed to the study design, obtaining funding and cohort collection. All authors have edited the manuscript and approved the final version.

**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 (80-. ).*, **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. Tafrechi, R.S.J., Svensson, J.P., Janssen, G.M.C., Szuhai, K., Maassen, J.A. and Raap, A.K. (2005) Distinct nuclear gene expression profiles in cells with mtDNA depletion and homoplasmic A3243G mutation. *Mutat. Res.*, **578**, 43–52.

9. Woodson, J.D. and Chory, J. (2008) Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.*, **9**, 383–395.

10. Chinnery, P.F., Elliott, H.R., Patel, S., Lambert, C., Keers, S.M., Durham, S.E., McCarthy, M.I., Hitman, G.A., Hattersley, A.T. and Walker, M. (2005) Role of the mitochondrial DNA 16184-16193 poly-C tract in type 2 diabetes. *Lancet*, **366**, 1650–1651.

11. Park, K.S., Chan, J.C., Chuang, L.M., Suzuki, S., Araki, E., Nanjo, K., Ji, L., Ng, M., Nishi, M., Furuta, H., *et al.* (2008) A mitochondrial DNA variant at position 16189 is associated with type 2 diabetes mellitus in Asians. *Diabetologia*, **51**, 602–608.

12. Saxena, R., De Bakker, P.I.W., Singer, K., Mootha, V., Burtt, N., Hirschhorn, J.N., Gaudet, D., Isomaa, B., Daly, M.J., Groop, L., *et al.* (2006) Comprehensive Association Testing of Common Mitochondrial DNA Variation in Metabolic Disease. *Am. J. Hum. Genet.*, **79**, 54–61.

13. Hwang, S., Kwak, S.H., Bhak, J., Kang, H.S., Lee, Y.R., Koo, B.K., Park, K.S., Lee, H.K., 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.

15. 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.

16. Smyth, G.K. (2005) Limma: linear models for microarray data. In *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer, pp. 397–420.

17. 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.

18. 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.

19. Sulonen, A.-M., Ellonen, P., Almusa, H., Lepist, 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.

20. 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.

21. Van Oven, M. and Kayser, M. (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.*, **30**, E394.

22. 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.

23. 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.

24. Landgraf, A.J. and Lee, Y. (2015) Dimensionality reduction for binary data through the projection of natural parameters. *arXiv Prepr. arXiv1510.06112*.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

30. Mueller, E.E., Eder, W., Ebner, S., Schwaiger, E., Santic, D., Kreindl, T., Stanger, O., Paulweber, B., Iglseder, B., Oberkofler, H., *et al.* (2011) The Mitochondrial T16189C Polymorphism Is Associated with Coronary Artery Disease in Middle European Populations. *PLoS One*, **6**, e16455.

31. 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.

32. Ferrannini, E., Gastaldelli, A. and Iozzo, P. (2011) Pathophysiology of Prediabetes. *Med. Clin. North Am.*, **95**, 327–339.

33. Powell, E.E. and Kroon, P.A. (1994) Low density lipoprotein receptor and 3‐ hydroxy‐3‐methylglutaryl coenzyme A reductase gene expression in human mononuclear leukocytes is regulated coordinately and parallels gene expression in human liver. *J. Clin. Invest.*, **93**, 2168–2174.

34. Ziegler-Heitbrock, H.W.L. (2000) Definition of human blood monocytes. *J. Leukoc. Biol.*, **67**, 603–606.

35. 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**, 1822–30.

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 mtSNP.

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

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

Table 1. 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 |

Table 2. 52 probe-mtSNP associations which had p < 1.28×10-8. Beta-coefficient represents the proportion of 1 SD change in normalized gene expression intensity (mean = 0, SD = 1). Asterisk (\*) marks the associations that were replicated from the results by Kassam et al. (7).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Gene | Illumina array ID | mtSNP | MAF | Beta | SE | P-value | Replication |
| SPCS2P4 | 4210315 | G9055A | 0.04 | -1.11 | 0.08 | 2.06 x 10-39 |  |
| SPCS2P4 | 4210315 | A3480G | 0.04 | -1.08 | 0.09 | 5.90 x 10-32 | \* |
| SPCS2P4 | 4210315 | C14167T | 0.04 | -1.08 | 0.09 | 1.29 x 10-31 |  |
| SPCS2P4 | 4210315 | A10550G | 0.04 | -1.08 | 0.09 | 1.32 x 10-31 | \* |
| SPCS2P4 | 4210315 | T16224C | 0.05 | -0.95 | 0.08 | 2.68 x 10-30 |  |
| SPCS2P4 | 4210315 | T11299C | 0.04 | -1.03 | 0.09 | 7.41 x 10-29 |  |
| SPCS2P4 | 4210315 | T1189C | 0.03 | -1.03 | 0.09 | 1.90 x 10-27 |  |
| RNF113A | 6940196 | G9055A | 0.04 | -1.23 | 0.13 | 2.26 x 10-21 |  |
| SPCS2P4 | 4210315 | T9698C | 0.05 | -0.73 | 0.08 | 1.93 x 10-19 | \* |
| RNF113A | 6940196 | C14167T | 0.04 | -1.25 | 0.14 | 1.67 x 10-18 |  |
| RNF113A | 6940196 | A10550G | 0.04 | -1.24 | 0.14 | 1.76 x 10-18 |  |
| RNF113A | 6940196 | A3480G | 0.04 | -1.24 | 0.14 | 2.37 x 10-18 |  |
| SPCS2P4 | 4210315 | A9093G | 0.02 | -1.02 | 0.12 | 5.90 x 10-18 |  |
| RNF113A | 6940196 | T11299C | 0.04 | -1.20 | 0.14 | 1.04 x 10-17 |  |
| RNF113A | 6940196 | T1189C | 0.03 | -1.23 | 0.14 | 4.06 x 10-17 |  |
| SPCS2P4 | 4210315 | T9903C | 0.02 | -1.06 | 0.12 | 6.34 x 10-17 |  |
| SPCS2P4 | 4210315 | T14798C | 0.10 | -0.60 | 0.07 | 1.09 x 10-15 |  |
| RNF113A | 6940196 | T16224C | 0.05 | -0.96 | 0.13 | 7.67 x 10-14 |  |
| SPCS2 | 7040068 | G9055A | 0.04 | -0.92 | 0.12 | 1.34 x 10-13 |  |
| SPCS2P4 | 4210315 | A1811G | 0.08 | -0.49 | 0.07 | 1.38 x 10-13 |  |
| CRLS1 | 2710446 | A8869G | 0.02 | -0.76 | 0.10 | 2.96 x 10-13 |  |
| CRLS1 | 2710446 | T4639C | 0.02 | -0.76 | 0.10 | 4.74 x 10-13 |  |
| RNF113A | 6940196 | T9698C | 0.05 | -0.87 | 0.12 | 2.73 x 10-12 |  |
| SPCS2P4 | 4210315 | G11377A | 0.03 | -0.75 | 0.11 | 3.12 x 10-12 |  |
| CRLS1 | 2710446 | C5263T | 0.02 | -0.76 | 0.11 | 3.61 x 10-12 |  |
| RNF113A | 6940196 | A9093G | 0.02 | -1.23 | 0.18 | 7.68 x 10-12 |  |
| MT-CO2 | 6550386 | G8269A | 0.01 | -1.63 | 0.24 | 1.37 x 10-11 | \* |
| SPCS2P4 | 4210315 | A11251G | 0.12 | 1.03 | 0.15 | 3.97 x 10-11 |  |
| SPCS2P4 | 4210315 | C15452A | 0.12 | 1.02 | 0.15 | 6.96 x 10-11 |  |
| RNF113A | 6940196 | T9903C | 0.02 | -1.24 | 0.19 | 1.95 x 10-10 |  |
| SLC25A14 | 1710754 | A3505G | 0.05 | 0.77 | 0.12 | 3.87 x 10-10 |  |
| RNF113A | 6940196 | T14798C | 0.10 | -0.72 | 0.11 | 3.88 x 10-10 |  |
| MT-RNR2L1 | 1230164 | C16256T | 0.07 | 0.39 | 0.06 | 4.59 x 10-10 |  |
| SLC25A14 | 1710754 | T1243C | 0.05 | 0.76 | 0.12 | 4.70 x 10-10 |  |
| SPCS2 | 7040068 | A3480G | 0.04 | -0.85 | 0.14 | 7.00 x 10-10 | \* |
| SPCS2 | 7040068 | A10550G | 0.04 | -0.85 | 0.14 | 7.53 x 10-10 | \* |
| SPCS2 | 7040068 | C14167T | 0.04 | -0.85 | 0.14 | 7.86 x 10-10 |  |
| SLC25A14 | 1710754 | A11947G | 0.05 | 0.74 | 0.12 | 1.10 x 10-9 |  |
| SLC25A14 | 1710754 | G8994A | 0.05 | 0.74 | 0.12 | 1.24 x 10-9 |  |
| RNF113A | 6940196 | T4216C | 0.12 | 1.32 | 0.22 | 1.44 x 10-9 |  |
| SPCS2 | 7040068 | T16224C | 0.05 | -0.75 | 0.12 | 1.70 x 10-9 |  |
| SPCS2P4 | 4210315 | A10398G | 0.14 | -0.40 | 0.07 | 1.79 x 10-9 |  |
| SLC25A14 | 1710754 | G5046A | 0.05 | 0.74 | 0.12 | 1.88 x 10-9 |  |
| SLC25A14 | 1710754 | G15884C | 0.05 | 0.74 | 0.12 | 2.49 x 10-9 |  |
| RNF113A | 6940196 | A1811G | 0.08 | -0.59 | 0.10 | 4.79 x 10-9 |  |
| RNF113A | 6940196 | C15452A | 0.12 | 1.39 | 0.24 | 4.98 x 10-9 |  |
| MT-CO2 | 6550386 | A16162G | 0.06 | -0.65 | 0.11 | 5.65 x 10-9 |  |
| RNF113A | 6940196 | A11251G | 0.12 | 1.36 | 0.23 | 7.86 x 10-9 |  |
| SLC25A14 | 1710754 | T12414C | 0.05 | 0.68 | 0.12 | 8.84 x 10-9 |  |
| SPCS2 | 7040068 | T11299C | 0.04 | -0.78 | 0.13 | 1.01 x 10-8 |  |
| SLC25A14 | 1710754 | G5460A | 0.05 | 0.65 | 0.11 | 1.16 x 10-8 |  |
| RNF113A | 6940196 | G11377A | 0.03 | -0.93 | 0.16 | 1.26 x 10-8 |  |

Abbreviations:MAF, minor allele frequency; 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 in 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 |
| *SPCS2P4* |  |  |
| 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 |
|  |  |  |
| *RNF113A* |  |  |
| K-H | -1.44 [-1.97, -0.91] |  |
| K-U | -1.53 [-2.07, -0.99] |  |
| K-J | -1.52 [-2.14, -0.90] | 2.48 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 |
|  |  |  |
| *SLC25A14* |  |  |
| 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 |
|  |  |  |
| *SLC2A8* |  |  |
| 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 |
|  |  |  |
| *MT-ND5* |  |  |
| 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

**Abbreviations**

mtDNA, mitochondrial DNA; mtSNP, mitochondrial single nucleotide polymorphism; T2DM, type 2 diabetes mellitus; T1DM, type 1 diabetes mellitus, 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