

Mitochondrial related genome-wide Mendelian randomization identifies putatively causal genes for multiple cancer types

Yanni Li,^{a,*} Kristina Sundquist,^{a,b,c,d} Naiqi Zhang,^a Xiao Wang,^a Jan Sundquist,^{a,b,c,d} and Ashfaq A. Memon^a

^aCenter for Primary Health Care Research, Lund University/Region Skåne, Malmö, Sweden

^bDepartment of Family Medicine and Community Health, Icahn School of Medicine at Mount Sinai, New York, United States

^cDepartment of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, United States

^dCenter for Community-Based Healthcare Research and Education (CoHRE), Department of Functional Pathology, School of Medicine, Shimane University, Matsue, Japan



Summary

Background Mitochondrial dysfunction is a hallmark of cancer. However, it is unclear whether it is a cause of cancer. This two-sample Mendelian randomization (MR) analyses, uses genetic instruments to proxy the exposure of mitochondrial dysfunction and cancer summary statistics as outcomes, allowing for causal inferences.

Methods Summary statistics from 18 common cancers (2107–491,974 participants), gene expression, DNA methylation and protein expression quantitative trait loci (eQTL, mQTL and pQTL, respectively, 1000–31,684 participants) on individuals of European ancestry, were included. Genetic variants located within or close to the 1136 mitochondrial-related genes (in *cis*) and robustly associated with the mitochondrial molecular alterations were used as instrumental variables, and their causal associations with cancers were examined using summary-data-based MR (SMR) analyses. An additional five MR methods were used as sensitivity analyses to confirm the casual associations. A Bayesian test for colocalization between mitochondrial molecular QTLs and cancer risk loci was performed to provide insights into the potential regulatory mechanisms of risk variants on cancers.

Findings We identified potential causal relationships between mitochondrial-related genes and breast, prostate, gastric, lung cancer and melanoma by primary SMR analyses. The sensitivity and the colocalization analyses further refined four genes that have causal effects on three types of cancer. We found strong evidence of positive association of *FDPS* expression level with breast cancer risk (OR per SD, 0.66; 95% CI, 0.49–0.83; $P = 9.77 \times 10^{-7}$), *NSUN4* expression level with both breast cancer risk (OR per SD, 1.05; 95% CI, 1.03–1.07; $P = 5.24 \times 10^{-6}$) and prostate cancer risk (OR per SD, 1.06; 95% CI, 1.03–1.09; $P = 1.01 \times 10^{-5}$), *NSUN4* methylation level with both breast and prostate cancer risk, and *VARS2* methylation level with lung cancer risk.

Interpretations This data-driven MR study demonstrated the causal role of mitochondrial dysfunction in multiple cancers. Furthermore, this study identified candidate genes that can be the targets of potential pharmacological agents for cancer prevention.

Funding This work was supported by Styrelsen för Allmänna Sjukhusets i Malmö Stiftelse för bekämpande av cancer (20211025).

Copyright © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Mendelian randomization; Mitochondrial dysfunction; Cancers; Colocalization; Pharmaceutical targets

Abbreviations: MR, Mendelian randomization; SMR, Summary-data-based MR; MtDNA, Mitochondrial DNA; IVs, Instrumental variables; GWAS, Genome-wide association studies; SNPs, Single nucleotide polymorphisms; eQTL, Expression quantitative trait loci; mQTL, Methylation quantitative trait loci; pQTL, Protein quantitative trait loci; HEIDI, Heterogeneity independent instruments; LD, Linkage disequilibrium; CpG, Cytosine-guanine dinucleotides; IVW, Inverse variance weighting; MR-PRESSO, MR Pleiotropy Residual Sum and Outlier

*Corresponding author. Center for Primary Health Care Research, Wallenberg Laboratory, Skåne University Hospital, 5th floor, Inga Marie Nilssons gata 53, S-205 02, Malmö, Sweden.

E-mail address: yanni.li@med.lu.se (Y. Li).

Research in context

Evidence before this study

Previous studies have shown associations between dysfunctions in mitochondrial DNA (mtDNA), mtDNA copy number or mitochondrial-related nuclear genes and different cancer risks. However, these studies did not investigate causal inferences between mitochondrial dysfunction and cancers. We searched PubMed for studies in any language using the search terms “mitochondrion OR mitochondria OR mitochondrial dysfunction” AND “Mendelian randomization OR Mendelian randomisation” AND “cancer OR cancers”. Of the yielded 4 studies, three studies’ outcomes were COVID-19, dementia and type 2 diabetes, respectively. The other study was a meta-analysis study that presented heterogeneous estimates for the effect of mtDNA copy numbers on different cancer risks and suggested applying Mendelian randomization for unraveling the casual correlation of mtDNA copy number with cancer risk.

Added value of this study

This data-driven study fills the gap by using Mendelian randomization to examine the potential causal relationship between mitochondrial dysfunction characterized by genetic predisposition in all mitochondrial-related genes and

common cancer risks. Our findings provide evidence for the potential causal effect of mitochondrial dysfunction on breast, prostate and lung cancer, after sensitivity and colocalization analyses. In addition, we identified a shared putative causal gene, *NSUN4*, for both breast and prostate cancer. All associations underscore the importance of mitochondrial dysfunction in the pathogenesis of multiple cancer types.

Implications of all the available evidence

Our data-driven analyses support the increasing values in the application of publicly accessible datasets to inform public health. To date, our European population-based large-scale study and the available evidence, indicate that individuals with mitochondrial dysfunction have a higher risk of a certain type of cancer, and point to the necessity of objective measurement of mitochondrial function in epidemiologic studies. For the identified putative causal genes, it is feasible to be added to the genetic screening project for better cancer prevention. True causal effects of mitochondrial dysfunction on cancers might be more complex and need larger genetic datasets and sophisticated experimental studies to further confirm.

Introduction

Mitochondria are the essential organelles that regulate cellular energy production, metabolism, proliferation and apoptosis. An altered mitochondrial function is a well-known hallmark of cancer, which is commonly characterized by abnormal mitochondrial morphology, deficient mitochondrial copy numbers, aberrant energetic metabolism, accumulation of reactive oxygen species (ROS), imbalanced biogenesis and mitophagy.¹ A mild mitochondrial dysfunction may enhance the amplification and invasion of cancer cells while a severe level of dysfunction may cause cell death to inhibit tumorigenesis. Thus, understanding the roles of mitochondrial dysfunction is essential for cancer research. Mitochondrial dysfunction is a complex cellular process that exhibits a spectrum of pathological conditions although there is no specific biomarker/s to define mitochondrial dysfunction.^{2,3} With the exception of the 37 critical bioenergetic genes encoded by the mitochondrion itself, the mitochondrial-related genome encompasses more than 1000 additional nuclear genes, and the genetic predisposition in those genes will potentially cause mitochondrial dysfunction.⁴ Many experimental and epidemiological studies have attempted to infer the causal relationship between mitochondrial dysfunction and cancer by exploring the selective mitochondrial DNA (mtDNA) and mitochondrial-related nuclear DNA mutations that affect mitochondrial function and are associated with the risk of specific cancer types.^{5,6} However, the results generated from

those studies are inconsistent and one of the reasons is the methodologies used in these studies, which do not consider the effect of confounders to differentiate between cause and consequence. Therefore, a comprehensive analysis of all genes related to mitochondrial dysfunction in multiple cancer types by a robust method is required to determine whether mitochondrial dysfunction per se is a cause or consequence of cancer.

Mendelian randomization (MR) is a method that uses genetic variants as instrumental variables (IVs) to explore the potential causal association between lifetime exposure and outcome. In MR, the use of the conceptional random allocation of alleles avoids bias from unobserved confounders such as lifestyle and environmental factors and the problem with reverse causality.⁷ The two-sample MR allows for the assessment of the IVs-exposure association and IVs-outcome association generated from different populations.⁸ Genome-wide association studies (GWAS) exploit the genetic associations with traits based on single nucleotide polymorphisms (SNPs) and integration of the GWAS data with gene expression and methylation GWAS have allowed for the identification of expression or methylation quantitative trait loci (eQTL or mQTL).^{9,10} A summary-data-based MR (SMR) has extended and developed the conception of MR that can utilize the independent GWAS summary statistics data and QTL data to prioritize potential causal genes from hits identified in GWAS.⁹ By applying this method followed by a heterogeneity independent instruments (HEIDI) test, the potential causal associations were distinguished from the widespread linkage disequilibrium (LD) in the genome.

To our knowledge, there has been no MR study investigating the potential causal relationship between mitochondrial dysfunction and the risk of common types of cancer. Therefore, in this study, we aimed to investigate the causal relationship between mitochondrial dysfunction characterized by genetic predisposition in mitochondrial-related genes and multiple cancer types by the comprehensive two-sample MR analysis.

Methods

This study was conducted following the reporting guideline of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE, [Supplementary STROBE-MR checklist table](#)).¹¹

Study design

[Fig. 1](#) summarizes the design of the present study and the workflow of the selection of genetic variants and analytical methods. To determine the mitochondrial dysfunction characterized by the genetic predisposition in the mitochondrial-related genome constituting from both mitochondrion and nuclear, we extracted the inventory of 1136 known mitochondrial-related genes from the human MitoCarta3.0 database.⁴

To generate eQTL instruments for mitochondrial genes, genetic variants located within 1000 kb on either side of the coding sequence (in *cis*) that are robustly associated with gene expression were extracted using eQTLs summary statistics obtained from the eQTLGen Consortium (<https://www.eqtlgen.org/cis-eqtl.html>). The eQTLGen Consortium contains information on 10,317 trait-associated single nucleotide polymorphisms (SNPs) from 31,684 individuals.¹² However, the eQTLGen did not include variants associated with the expression level of genes located on the X and Y chromosomes and mtDNA. From *cis*-eQTL, 662,968 SNPs associated with the expression of 1013 mitochondrial-related transcripts were selected. MR *cis*-mQTL instruments for genetic variants robustly associated with mitochondrial gene methylation were extracted using summary data from a meta-analysis of two cohorts ($n = 1980$).¹⁰ In total, 931,304 SNPs were selected corresponding to 2550 mitochondrial-related DNA methylation CpG sites. MR *cis*-pQTL instruments for genetic variants associated with the expression of mitochondrial-related proteins were selected from five proteome datasets,^{13–17} and 23 SNPs that were robustly associated with 23 mitochondrial-related protein expressions were selected. All SNPs included in the initial analysis had at least a suggestive $P_{\text{snp-mitodys}} < 5 \times 10^{-8}$.

GWAS summary statistics for cancer outcomes were obtained from publicly available databases. A total of 18 types of cancers were included. The details of all QTL and GWAS datasets for this study are presented in [Supplementary Table S1](#) and [Supplementary methods](#).

Statistical analysis

The main analyses involved three stages: primary SMR analyses, sensitivity analyses and colocalization analyses.

Mendelian randomization requires meeting three core assumptions ([Supplementary methods](#)). As an extension of the MR concept, SMR was developed to estimate the pleiotropic association between genetically determined traits (e.g., gene expression, DNA methylation, or protein abundance as exposure) and complex traits of interest (e.g., disease phenotype as outcome).⁹ To meet MR assumptions in our study, the causal association was calculated as:

$$\beta_{\text{mitodys-cancer}} = \beta_{\text{SNP-cancer}} / \beta_{\text{SNP-mitodys}}$$

$\beta_{\text{mitodys-cancer}}$ is calculated as the estimated effect size of mitochondrial dysfunction on cancer, where $\beta_{\text{SNP-mitodys}}$ is the estimated effect size of SNP on mitochondrial dysfunction (a genetic variant—exposure trait association) and $\beta_{\text{SNP-cancer}}$ is the estimated effect size of SNP on cancer (the same genetic variant—outcome trait association).

Here, we performed SMR using the Linux version 1.0.3 of SMR software in the command line using default options (<https://yanglab.westlake.edu.cn/software/smr/#Overview>). Odds ratio (OR) estimates of mitochondrial dysfunction on the risk of cancer were obtained as follows: $\text{OR}_{\text{mitodys-cancer}} = \exp(\beta_{\text{mitodys-cancer}})$, where OR is the odds ratio estimate per 1-ln increment in mitochondrial genome levels and exp is the base of the natural logarithm.

Sensitivity analyses were conducted after completing the primary SMR analyses with 5 additional MR methods, including MR Egger, weighted median, inverse variance weighting (IVW), simple mode and weighted mode by using the TwoSampleMR R package. Each of these methods calculates the estimates of the causal effect based on slightly different assumptions about the instrument validity and therefore provide robust evidence of our findings ([Supplementary methods](#)). All analyses in this part were performed using R software (version 4.1.2, www.r-project.org).

HEIDI test is one of the colocalization methods that use external reference to estimate the LD. To refine the results, we performed another Bayesian test for the colocalization of two traits using the coloc R package (<https://chr1swallace.github.io/coloc/>, version 5.1.0) to estimate the posterior probability of shared variants.¹⁸ For each leading SNP in the investigated cancer GWAS database, all SNPs within 100 kb up and downstream of the leading SNPs were retrieved for colocalization analysis to analyze the posterior probability of H4 (PP.H4), and PP.H4 > 0.8 is the well-applied cut-off for the evidence of colocalization of the GWAS and QTL association ([Supplementary methods](#)).

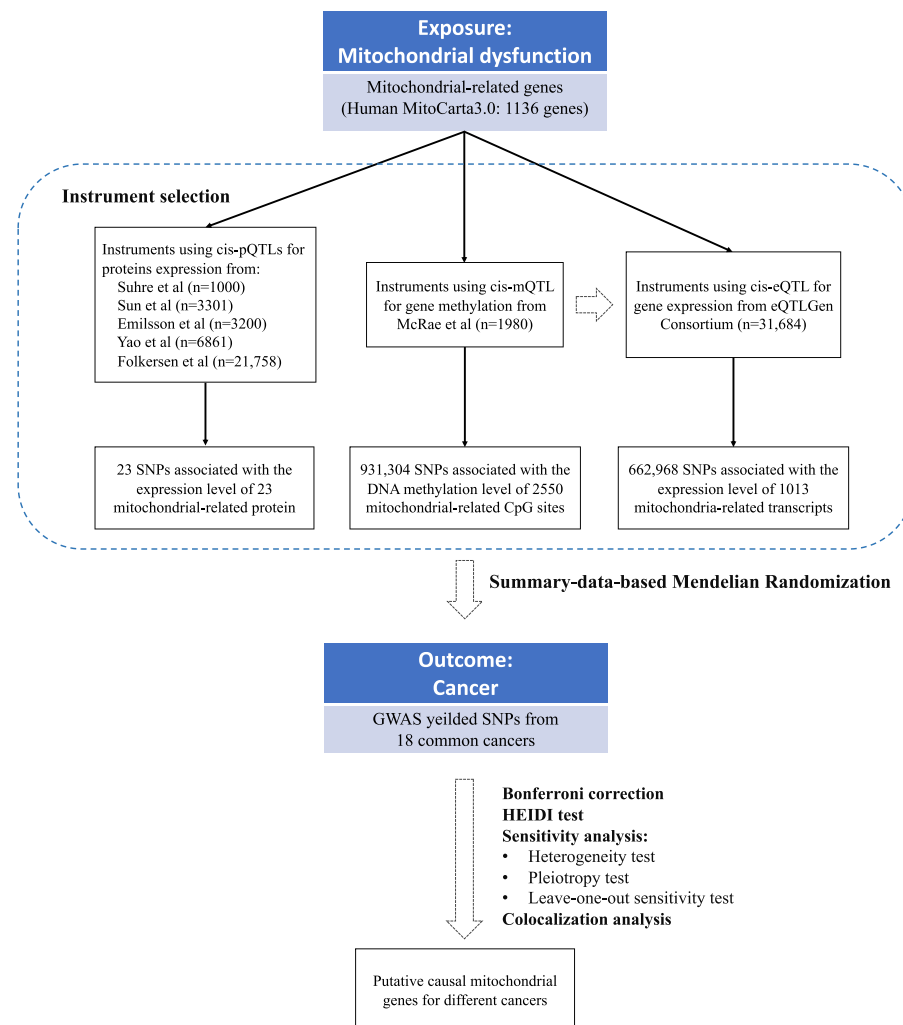


Fig. 1: Flowchart of the analyses performed.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Ethics

All summarized statistics utilized in the MR analyses were generated by previous studies, for which ethical approval and individual consent were obtained for all original studies.

Results

MR analysis of mitochondrial genome-wide cis-eQTLs and cancer outcomes

After SMR testing, the associations of 662,968 SNPs from blood representing mitochondrial-related gene

expression and cancer outcomes were obtained (Fig. 1). To control the genome-wide type I error, we performed multiple testing corrections, with the results showing strong evidence of an association ($P_{\text{SMR}} < 4.936 \times 10^{-5}$ [Bonferroni correction, $P < 0.05/1013$]) followed by the HEIDI test ($P_{\text{HEIDI}} > 0.01$) implemented in SMR software to investigate if the association was due to a shared causal variant and not pleiotropy. We thus identified 7 association signals across 7 unique genetic loci for breast cancer, 4 association signals across 4 unique genetic loci for prostate cancer and one association signal for gastric cancer. We found no significant genetic correlation for the other cancer types. Sensitivity analysis using additional MR methods relying on similar assumptions was conducted and shown to support our findings (Supplementary Table S2, Supplementary Figure S1 and S2). We further performed colocalization analysis to rule out confounding by LD; strong

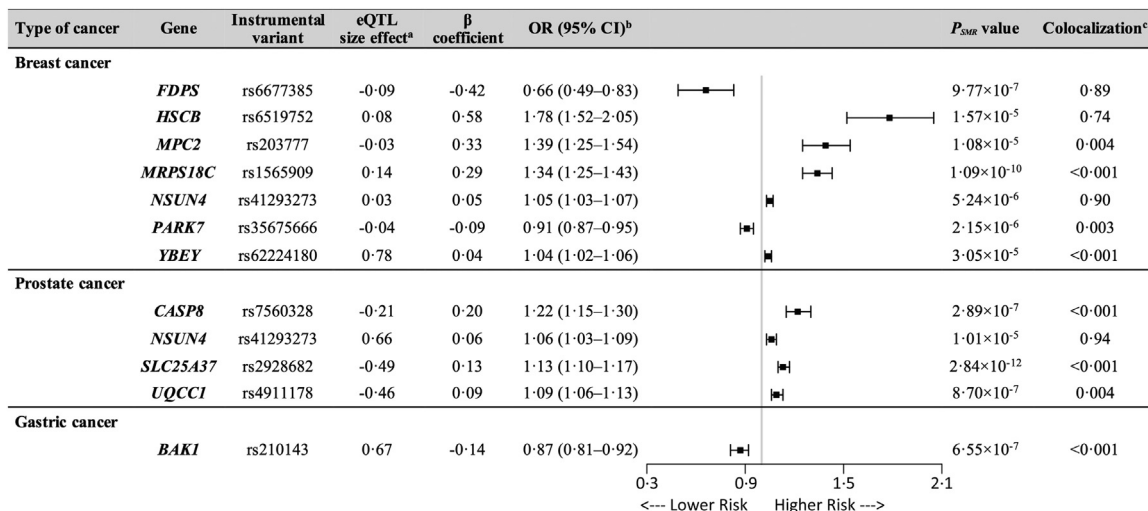


Fig. 2: Mendelian randomization results for the association between the expression of mitochondrial genes and cancer risk. ^aRepresents the effect size (β) of a variant on mRNA expressions. $\beta > 0$ means positive association, and $\beta < 0$ means negative association. ^bOdds ratios were calculated by the expectation of causal estimate (β coefficient). ^cColocalization indicates PP.H4 between eQTLs and cancer outcomes. PP.H4 > 0.8 is the well-applied cut-off for the evidence of colocalization.

evidence of colocalization between cancer GWAS and eQTL exists if the posterior probability of shared causal variant across gene expression and cancer (PP.H4) is >0.80. The causal estimates are expressed as β coefficients, and the odds ratios (OR) for 1 standard deviation (SD) change in mitochondrial gene expression level was calculated by the expectation of the β coefficient, as presented in Fig. 2 and Supplementary Table S3.

For breast cancer, one SD decrease of *FDPS* expression was associated with 34% lower risk (OR: 0.66, 95% CI: 0.49–0.83, $P_{SMR} = 9.77 \times 10^{-7}$) while 1 SD increase of *NSUN4* expression was associated with 5% higher risk (OR: 1.05, 95% CI: 1.03–1.07, $P_{SMR} = 5.24 \times 10^{-6}$). When sub-grouping breast cancer according to the intrinsic molecular subtypes, the causal associations showed a similar trend only with luminal A-like breast cancer. Interestingly, we found a robust causal association (OR per SD, 1.17; 95% CI: 1.12–1.23, $P_{SMR} = 1.85 \times 10^{-8}$) between *MTX1* expression and luminal A-like breast cancer, specifically. We also found a strong causal association (OR per SD, 1.26; 95% CI: 1.15–1.37, $P_{SMR} = 2.94 \times 10^{-5}$) between *COX11* expression and luminal B-like/HER2-negative cancer (Supplementary Table S4, Supplementary Fig. S3). For prostate cancer, one SD increase of *NSUN4* expression was associated with 6% higher risk of cancer (OR: 1.06, 95% CI: 1.03–1.09, $P_{SMR} = 1.01 \times 10^{-5}$).

Most importantly, our results show that the expression level of *NSUN4* increased by rs41293273 is associated with higher risk of both breast cancer and prostate cancer.

MR analysis of mitochondrial genome-wide cis-mQTLs and cancer outcomes

For the causal association between the DNA methylation of the mitochondrial-related genome and cancer outcomes, Bonferroni correction ($P_{SMR} < 1.961 \times 10^{-5}$) and HEIDI test were performed. We identified a total of 15 association signals across 14 unique genetic loci for breast cancer, 11 association signals across 10 unique genetic loci for prostate cancer, one association signal for gastric cancer, 4 association signals across 3 unique genetic loci for lung cancer and 2 association signals across one unique genetic locus for melanoma (Fig. 3 and Supplementary Table S5). The sensitivity analysis supported the same associations (Supplementary Table S6, Supplementary Fig. S4 and S5).

The colocalization analysis showed that different genetic variants regulating *NSUN4* had different effects on methylation levels, hence the outcome. For example, one SD decrease of *NSUN4* methylation by rs6682266 was associated with 9% lower risk of breast cancer (OR: 0.91, 95% CI: 0.87–0.96, $P_{SMR} = 1.50 \times 10^{-5}$), and conversely, one SD increase of *NSUN4* methylation by rs6681857 was associated with 7% higher risk of breast cancer (OR: 1.07, 95% CI: 1.04–1.10, $P_{SMR} = 5.81 \times 10^{-6}$). Here, in total, we found 6 unique loci that regulated the methylation level of 7 different CpG sites in *NUSU4*, and were positively associated with the risk of breast cancer (Fig. 3). Analysis on breast cancer molecular subtypes with *NSUN4* methylation showed a similar causal association but only for luminal A-like breast cancer (Supplementary Table S7 and Supplementary Fig. S6). Out of these 6 loci, two were

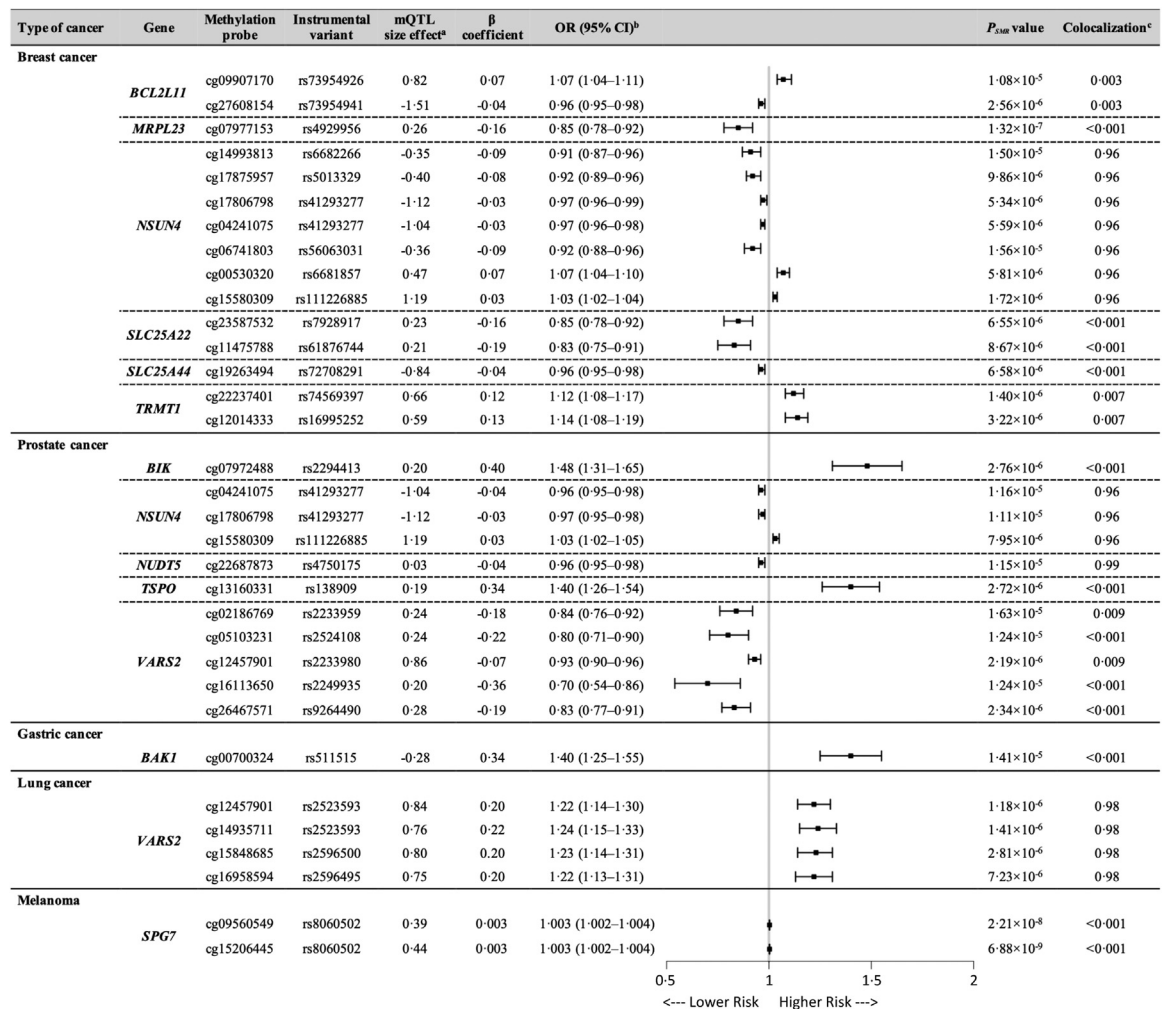


Fig. 3: Mendelian randomization results for the association between mitochondrial gene methylations and cancer risk. ^aRepresents the effect size (β) of a variant on DNA methylation. $\beta > 0$ means positive association, and $\beta < 0$ means negative association. ^bOdds ratios were calculated by the expectation of causal estimate (β coefficient). ^cColocalization indicates PP.H4 between mQTLs and cancer outcomes. PP.H4 > 0.8 is the well-applied cut-off for the evidence of colocalization.

also positively associated with risk of prostate cancer. Furthermore, an increase of 1 SD of *NUDT5* methylation was associated with 4% lower risk of prostate cancer (OR: 0.96, 95% CI: 0.95–0.98, $P_{SMR} = 1.15 \times 10^{-5}$) (Fig. 3). For lung cancer, one SD increase of *VARS2* methylation was associated with more than 20% higher risk of cancer depending on different methylation CpG sites (Fig. 3). We found no significant causal associations between mitochondrial genome methylation and other cancer risks.

Furthermore, gene methylation is known to influence gene expression. Here, we also performed SMR analysis on the causal association between mitochondrial-related gene methylation and expression by mapping the gene methylation to expression through shared genetic variants. After multiple testing

corrections and the HEIDI test, we obtained the gene list for the mitochondrial gene expression regulated by DNA methylation CpG sites (Supplementary Table S8). For the putative causal genes that we identified above, SMR results showed that the methylation of *NSUN4*, which was regulated by rs6682266, rs5013329, rs56063031 and rs6681857, was associated with *NSUN4* expression, and *VARS2* methylation by rs2596495 was also associated with *VARS2* expression (Supplementary Table S8).

MR analysis of mitochondrial genome-wide cis-pQTLs and cancer outcomes

Only 23 proposed mitochondrial-related SNPs were extracted from cis-pQTLs and no causal association was

found based on our suggested threshold after the SMR analyses. One possible explanation could be that the pQTL datasets are not comprehensively developed, and very few genetic variants that are robustly associated with protein levels were identified.

Phenome-wide scan of identified genetic variants

To exclude possible pleiotropy of investigated cancers, we performed phenome-wide scan analysis on the identified variants, using both GWASATLAS¹⁹ and PhenoScanner²⁰ databases. The databases enable the investigation of genetic variants across multiple disease traits. The Manhattan plots were used to show phenome-wide scan results of identified genetic variants on different disease traits with any possible effect allele obtained from GWASATLAS (Supplementary Fig. S8–S10). The phenome-wide scan of all identified genetic variants with disease traits from PhenoScanner was also listed according to the following selection criteria: 1) the SNPs shared the same effect allele with our results, 2) the association reached GWAS significance ($P < 5 \times 10^{-8}$) and 3) the absolute value of size effect (β) > 0.01 (Supplementary Table S9). Interestingly, variants related to the gene expression and DNA methylation of *NSUN4* that were causally associated with both breast and prostate cancer were not found to be associated with all available secondary traits (Supplementary Fig. S7 and S8). This further suggests that the causal relationship between *NSUN4* and breast and prostate cancer identified in this study is robust. However, rs6677385 (*FDPS* expression associated) was associated with secondary traits such as metabolic-related blood urea nitrogen and Crohn's disease (Supplementary Fig. S9a). The rs4750175 (*NUDT5* methylation associated) was also associated with endocrine-related traits (risk of type 2 diabetes, Supplementary Fig. S9b). Three SNPs associated with *VAR2* methylation were further associated with multiple traits such as metabolic, skeletal, respiratory, psychiatric, immunological and endocrine-related traits (Supplementary Fig. S10). The genetic variants that are associated with the secondary traits may potentially introduce horizontal pleiotropy, further investigations to rule out pleiotropy are needed.

Bi-directional MR analysis of mitochondrial dysfunction and cancers

GWAS summary statistics were available only for mtDNA copy number variation, which has been suggested as a surrogate biomarker for mitochondrial dysfunction. Currently, a GWAS dataset specifically containing genetic variant association with mtDNA copy number has been published.²¹ We used this dataset to explore whether mitochondrial dysfunction is a consequence of cancer and conducted bidirectional MR analyses on mtDNA copy number and cancers. Results

showed that the directions of causal association were cancer type-specific; here mtDNA copy number variation has causal effects on cervical cancer, specific subtype of ovarian cancer (Supplementary Table S10), while triple-negative breast cancer, head and neck cancer were causally associated with mtDNA copy number variation (Supplementary Table S11).

Discussion

In this study, we demonstrate that mitochondrial dysfunction characterized by genetic predisposition has causal effect on cancers, and identified important putative causal mitochondrial-related genes as follows: 1) *FDPS* for breast cancer; 2) *NUDT5* for prostate cancer; 3) *VAR2* for lung cancer and 4) *NSUN4* for both breast and prostate cancers. Our results show that genetic determinants of mitochondrial dysfunction were associated with the risk of cancer in a cancer type-specific manner, which provides robust evidence for underlying mechanisms linking the genetic loci, gene expression, and methylation with multiple cancers.

The *FDPS* is a key enzyme that is involved in the mevalonate pathway to catalyze the biosynthesis of cholesterol and sterol, and to isoprenylate cellular metabolites such as Ras, Rac, Rab and Rho for membrane anchorage and cellular signaling.²² *FDPS* has been investigated over decades for its physiological function and was found to be associated with leukemia growth,²³ the progression of prostate cancer,²⁴ the poor breast cancer prognosis,²⁵ and directly involved in glioblastoma drug resistance²⁶ and pancreatic cancer radioresistance.²⁷ However, its causal relationship with cancer is unclear. In this study, we show that gene expression of *FDPS* has a causal relationship with breast cancer. Studies have shown that knockdown *FDPS* enhanced apoptosis and ectopic overexpression of *FDPS* promoted cancer colony growth and proliferation by affecting STAT3, AKT and ERK pathways.²⁴ Prenylation is important for exerting the activity of oncogenic proteins, thus prenylation inhibitors have been widely applied in clinical trials for cancer treatment.²⁸ *FDPS* was shown to be a key target of nitrogen-containing bisphosphonates and it is already a clinical drug target by Zoledronic acid.²⁹ According to the DRUGBANK database (<https://go.drugbank.com/>), more drugs have been investigated such as Ibandronate, Minodronic acid and Incadronic acid that target *FDPS*.

NSUN4 is an rRNA m⁵C methyltransferase that can induce the methylation of the 12S rRNA of the small ribosomal subunit joining in mitochondria and promote rRNA rearrangements to form peptidyl transferase center.^{30,31} A previous study showed that breast cancer and prostate cancer shared a common risk locus (rs5013329) and indicated that *NSUN4* is the strongest shared functional candidate at 1p34.³² However, our results showed that rs5013329 related to *NSUN4*

methylation was associated with decreased risk of breast cancer only. Importantly, we identified additional 1 genetic locus related to *NSUN4* expression and 2 loci related to *NSUN4* methylation that were causally associated with both breast and prostate cancer (Supplementary Fig. S7). Furthermore, our phenome-wide scan analysis showed that the causal relationship between *NSUN4* and both breast and prostate cancers was not caused by horizontal pleiotropy. Together, these results emphasize the potentially important role of *NSUN4* in carcinogenesis.

NUDT5 is differentially expressed in different types of cancer and positively correlated with aggressive cancer disease phenotype, knockdown of which can suppress the proliferation of cancer cells without inducing DNA oxidative lesion.³³ In our study, we show that the methylation level of *NUDT5* has a strong causal effect on prostate cancer. Those findings highlight that *NUDT5* may represent a promising drug target for cancer prevention and treatment. More studies now focus on the identification of *NUDT5* inhibitors from approved drugs and small molecules, and a potent TH5427 was tested and shown to block hormone signaling and disrupt the proliferation of breast cancer cells.³⁴

Several mutations in *VAR2* have been associated with mitochondrial diseases such as complex I defect, early onset of mitochondrial encephalomyopathies and encephalocardiomyopathies,³⁵ and cancer risks including breast cancer, colon and lung cancer.^{36–38} Here, we propose *VAR2* as a causal gene only for lung cancer. The mechanisms of *VAR2* in lung cancer carcinogenesis need further evaluation by experimental studies.

The main strength of the present study is that we performed a comprehensive MR analysis between mitochondrial dysfunction, characterized by genetic predisposition in all known mitochondrial-related genes, and their causal relationship with cancers. The inclusion of all genes related to mitochondria eliminates the selection bias in previous studies and might be able to address mitochondrial dysfunction directly. Secondly, we have included a very large sample size and 18 different cancer outcomes from GWAS summarized statistics, which allowed us to gain sufficient power to elucidate causal relationships and make conclusive estimations for several cancer types. Thirdly, we used SMR as the primary analysis and performed a sensitivity analysis using 5 additional MR approaches and colocalization analysis, which shows the robustness of our findings. Finally, we only included samples of European ancestry, thus, we minimized the biases caused by different genetic backgrounds.

This study has several limitations as well. Although we drew on the large available GWAS data sources, no genetic variants were obtained that represent the mitochondrial protein expression and the available eQTL and

mQTL datasets did not have information on genetic variants that were associated with gene expression or methylation level in the X chromosome, Y chromosome and mitochondrial genome; the mitochondrial genome-wide associated genetic variants in this study mainly laid on the mitochondrial-related nuclear genome rather than the mitochondrial genome itself because a mitochondrial genome-specific QTL dataset has not yet been developed. Moreover, GWAS dataset that directly reflected on mitochondrial dysfunction is not available, hence we cannot assess the direction of causal relationship by using bi-directional MR based on current software resources. In this study, we showed that causal effects of mtDNA copy number variations and cancers were bi-directional in a cancer-specific manner. However, GWAS summary statistics of mtDNA copy number variation are potentially underpowered to detect the direction of causal association between mitochondrial dysfunction and cancers. Furthermore, univariable MR estimates the total effect of exposure on the outcome. As an extension, multivariable MR simultaneously estimates several potentially related exposures with a shared set of SNPs on the outcome using GWAS summary statistics, allowing for the assessment of the direct causal effect of a single exposure on the outcome. In this study, the exposure was mitochondrial dysfunction characterized by predisposition in the mitochondrial-related gene, which only can be retrieved from the QTL datasets other than the GWAS dataset. Thus, we are unable to perform multivariable MR to estimate the direct causal effects of mitochondrial dysfunction on cancers. Further studies should be conducted on the question of whether mitochondrial dysfunction is causally associated with cancer when GWAS or more advanced methods are available.

This study leverages MR to examine the potential causal relationship between mitochondrial dysfunction characterized by genetic predisposition in mitochondrial-related genes and cancer, and demonstrates the importance of mitochondrial dysfunction in the pathogenesis of multiple cancer types. The identified putative genes can function as potential pharmacological targets for cancer treatment and prevention, further research could explore details of the underlying biological mechanisms.

Contributors

YL designed the study and performed the statistical analysis; YL, KS, NZ, XW and AAM participated in data interpretation; YL and AAM wrote the first draft, and KS, NZ, XW and JS revised the article. YL and NZ had direct access and responsibility for verifying all data reported in the manuscript. All authors read and approved the submitted version of the manuscript.

Data sharing statement

Data used in this study are available from the referenced peer-reviewed studies and listed in Supplementary Table S1. Summary statistics for

GWAS are publicly available for download. The statistical code needed to reproduce the results in the article is available upon request.

Declaration of interests

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

Acknowledgments

We would like to thank Patrick O'Reilly for proofreading the article. This work was supported by Styrelsen för Allmänna Sjukhusets i Malmö Stiftelse för bekämpande av cancer (20211025).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2022.104432>.

References

- Bao X, Zhang J, Huang G, et al. The crosstalk between HIFs and mitochondrial dysfunctions in cancer development. *Cell Death Dis.* 2021;12(2):215.
- Diaz-Vegas A, Sanchez-Aguilera P, Krycer JR, et al. Is mitochondrial dysfunction a common root of noncommunicable chronic diseases? *Endocr Rev.* 2020;41(3):bnaa005.
- Miwa S, Kashyap S, Chini E, von Zglinicki T. Mitochondrial dysfunction in cell senescence and aging. *J Clin Invest.* 2022;132(13):e158447.
- Rath S, Sharma R, Gupta R, et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res.* 2021;49(D1):D1541–D1547.
- Peduzzi G, Gentiluomo M, Tavano F, et al. Genetic polymorphisms involved in mitochondrial metabolism and pancreatic cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2021;30(12):2342–2345.
- Moro L. Mitochondrial DNA and MitomiR variations in pancreatic cancer: potential diagnostic and prognostic biomarkers. *Int J Mol Sci.* 2021;22(18).
- Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ.* 2018;362:k601.
- Lawlor DA. Commentary: two-sample Mendelian randomization: opportunities and challenges. *Int J Epidemiol.* 2016;45(3):908–915.
- Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48(5):481–487.
- McRae AF, Marioni RE, Shah S, et al. Identification of 55,000 replicated DNA methylation QTL. *Sci Rep.* 2018;8(1):17605.
- Cuscheri V. The STROBE guidelines. *Saudi J Anaesth.* 2019;13(Suppl 1):S31–S34.
- Vosa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet.* 2021;53(9):1300–1310.
- Ahmad A, Sundquist K, Zoller B, et al. Evaluation of expression level of apolipoprotein M as a diagnostic marker for primary venous thromboembolism. *Clin Appl Thromb Hemost.* 2018;24(3):416–422.
- Emilsson V, Ilkov M, Lamb JR, et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science.* 2018;361(6404):769–773.
- Suhre K, Arnold M, Bhagwat AM, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun.* 2017;8:14357.
- Yao C, Chen G, Song C, et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat Commun.* 2018;9(1):3268.
- Folkersen L, Gustafsson S, Wang Q, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab.* 2020;2(10):1135–1148.
- Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10(5):e1004383.
- Watanabe K, Stringer S, Frei O, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet.* 2019;51(9):1339–1348.
- Kamat MA, Blackshaw JA, Young R, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics.* 2019;35(22):4851–4853.
- Pillalamarri V, Shi W, Say C, et al. Whole-exome sequencing in 415,422 individuals identifies rare variants associated with mitochondrial DNA copy number. *HGG Adv.* 2022;4(1):100147.
- Sun S, McKenna CE. Farnesyl pyrophosphate synthase modulators: a patent review (2006 - 2010). *Expert Opin Ther Pat.* 2011;21(9):1433–1451.
- Zhang F, Dai X, Wang Y. 5-Aza-2'-deoxycytidine induced growth inhibition of leukemia cells through modulating endogenous cholesterol biosynthesis. *Mol Cell Proteomics.* 2012;11(7):M111.016915.
- Seshacharyulu P, Rachagani S, Muniyan S, et al. FDNs cooperates with PTEN loss to promote prostate cancer progression through modulation of small GTPases/AKT axis. *Oncogene.* 2019;38(26):5265–5280.
- Clendening JW, Pandya A, Boutros PC, et al. Dysregulation of the mevalonate pathway promotes transformation. *Proc Natl Acad Sci U S A.* 2010;107(34):15051–15056.
- Woo IS, Eun SY, Kim HJ, et al. Farnesyl diphosphate synthase attenuates paclitaxel-induced apoptotic cell death in human glioblastoma U87MG cells. *Neurosci Lett.* 2010;474(2):115–120.
- Souček JJ, Baine MJ, Lin C, et al. Unbiased analysis of pancreatic cancer radiation resistance reveals cholesterol biosynthesis as a novel target for radiosensitisation. *Br J Cancer.* 2014;111(6):1139–1149.
- Berndt N, Hamilton AD, Sebt SM. Targeting protein prenylation for cancer therapy. *Nat Rev Cancer.* 2011;11(11):775–791.
- Gritzalis D, Park J, Chiu W, et al. Probing the molecular and structural elements of ligands binding to the active site versus an allosteric pocket of the human farnesyl pyrophosphate synthase. *Bioorg Med Chem Lett.* 2015;25(5):1117–1123.
- Camara Y, Asin-Cayuela J, Park CB, et al. MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome. *Cell Metab.* 2011;13(5):527–539.
- Spahr H, Habermann B, Gustafsson CM, Larsson NG, Hallberg BM. Structure of the human MTERF4-NSUN4 protein complex that regulates mitochondrial ribosome biogenesis. *Proc Natl Acad Sci U S A.* 2012;109(38):15253–15258.
- Kar SP, Beesley J, Amin Al Olama A, et al. Genome-wide meta-analyses of breast, ovarian, and prostate cancer association studies identify multiple new susceptibility loci shared by at least two cancer types. *Cancer Discov.* 2016;6(9):1052–1067.
- Pickup KE, Pardow F, Carbonell-Caballero J, et al. Expression of oncogenic drivers in 3D cell culture depends on nuclear ATP synthesis by NUDT5. *Cancers.* 2019;11(9):1337.
- Tong XY, Liao X, Gao M, et al. Identification of NUDT5 inhibitors from approved drugs. *Front Mol Biosci.* 2020;7:44.
- Bruni F, Di Meo I, Bellacchio E, et al. Clinical, biochemical, and genetic features associated with VARS2-related mitochondrial disease. *Hum Mutat.* 2018;39(4):563–578.
- Chae YS, Lee SJ, Moon JH, et al. VARS2 V552V variant as prognostic marker in patients with early breast cancer. *Med Oncol.* 2011;28(4):1273–1280.
- Zhu Z, Hou Q, Wang B, et al. A novel mitochondria-related gene signature for controlling colon cancer cell mitochondrial respiration and proliferation. *Hum Cell.* 2022;35(4):1126–1139.
- van Kooij M, de Groot K, van Vugt H, Aten J, Snoek M. Genotype versus phenotype: conflicting results in mapping a lung tumor susceptibility locus to the G7c recombination interval in the mouse MHC class III region. *Immunogenetics.* 2001;53(8):656–661.