LUND UNIVERSITY

Role of mitochondria in early molecular diagnosis and prognosis of cancer

Li, Yanni

2023

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

Link to publication

Citation for published version (APA):
Li, Y. (2023). *Role of mitochondria in early molecular diagnosis and prognosis of cancer*. [Doctoral Thesis (compilation), Department of Clinical Sciences, Malmö]. Lund University, Faculty of Medicine.

Total number of authors:
1

Creative Commons License:
Unspecified

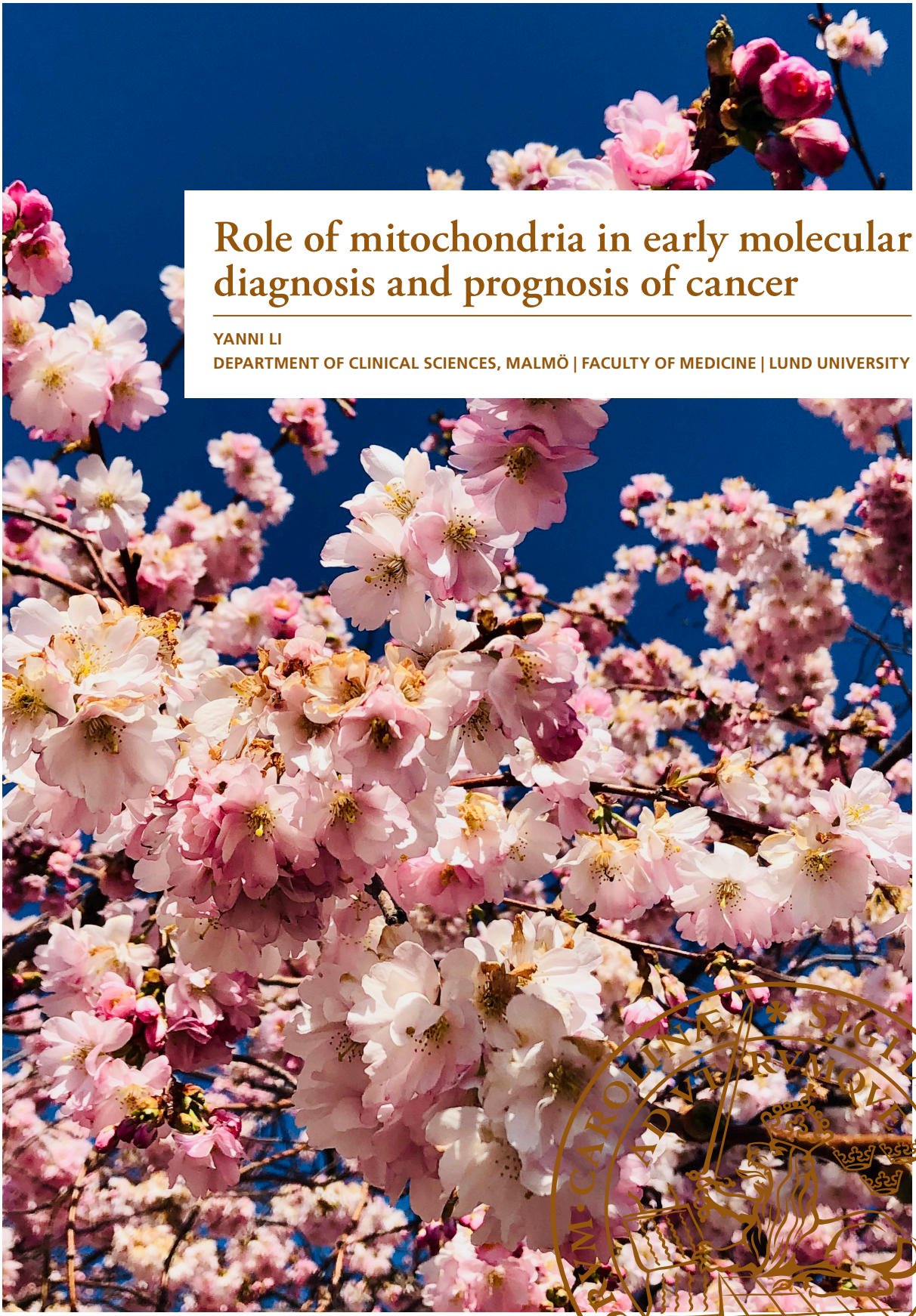
General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



Role of mitochondria in early molecular diagnosis and prognosis of cancer

YANNI LI
DEPARTMENT OF CLINICAL SCIENCES, MALMÖ | FACULTY OF MEDICINE | LUND UNIVERSITY

Role of mitochondria in early molecular diagnosis and prognosis of cancer

Role of mitochondria in early molecular diagnosis and prognosis of cancer

Yanni Li

李艳妮

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University.

To be publicly defended on 29th of September 2023 at 13:00 in Agardh Hall, CRC, Department of Clinical Science, Jan Waldenströms gata 35, Malmö

Faculty opponent
Xuefeng Zhu, Associate Professor

Department of Medical Biochemistry and Cell Biology,

University of Gothenburg

Organization: LUND UNIVERSITY

Document name: Doctoral Dissertation **Date of issue:** 2023-09-29

Author(s): Yanni Li **Sponsoring organization**

Title and subtitle: Role of mitochondria in early molecular diagnosis and prognosis of cancer

Abstract:

Background:

Earlier clinical detection of cancer may improve survival as well as offer opportunities for less invasive treatment options. This thesis explores whether the mitochondria and its related genes in the nuclear genome can be used as novel methods for the diagnosis and prognosis of cancers.

Aims and Methods:

Paper I: To investigate if mitochondrial dysfunction (characterized by mtDNA copy number variations) is associated with prevalent, incident cancer and cancer mortality – droplet digital PCR (ddPCR).
Paper II: To investigate the potential causal relationship between mitochondrial dysfunction (characterized by genetic predispositions in all mitochondrial-related genes) and common cancer risks – Mendelian randomization, colocalization.
Paper III: To investigate mitochondrial mutations as potential biomarkers for the early diagnosis of breast cancer – whole mitochondrial genome sequencing, bioinformatics, ddPCR.
Paper IV: To investigate the mitochondrial-related gene expression signature as a prognostic model to predict the clinical outcome for breast cancer patients – machine learning.

Results and conclusions:

Paper I: We found that mtDNA-CN was significantly associated with prevalent and incident cancer as well as cancer mortality. However, these associations were cancer-type specific and need further investigation.
Paper II: We identified potential causal relationships between mitochondrial-related genes and breast, prostate and lung cancer. Furthermore, this study identified candidate genes that can be the targets of potential pharmacological agents for cancer prevention.
Paper III: We comprehensively characterized the mtDNA mutation landscape of breast cancer biopsies and matched baseline whole blood samples. Notably, we have identified and validated mt.16093T>C mutation, which was associated with a 67% increased risk of developing breast cancer, and could potentially be used as early breast cancer diagnostic biomarkers.
Paper IV: We built a novel 14 genes mitochondrial signature model that could be an independent prognostic predictor and together with clinical variables as an improved model for predicting overall early-stage of breast cancer survival.

Key words: cancer, mtDNA copy number, mitochondrial-related nuclear genes, mtDNA mutations, diagnosis, prognosis, biomarker

Classification system and/or index terms (if any) Supplementary bibliographical information

Language: English **ISSN and key title**: 1652-8220

ISBN: 978-91-8021-454-4

Recipient's notes **Number of pages**: 86

Price Security classification

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature 李艳妮 Yanni Li Date 2023-08-17

Role of mitochondria in early molecular diagnosis and prognosis of cancer

Yanni Li

李艳妮

LUND
UNIVERSITY

Coverphoto by Yanni Li
Copyright pp 1-86 Yanni Li

Paper 1 © Li et al, 2021
Paper 2 © Li et al, 2023
Paper 3 © by the Authors (Manuscript unpublished)
Paper 4 © by the Authors (Manuscript unpublished)

Faculty of Medicine
Department of Clinical Sciences, Malmö

ISBN 978-91-8021-454-4
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2022

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se
MADE IN SWEDEN

The way ahead is long and has no ending,
Yet high and low, I will search with my will unbending.
-- Qu Yuan

路漫漫其修远兮，吾将上下而求索
-- 屈原

Table of Contents

Abstract

Background:

Earlier clinical detection of cancer may improve survival as well as offer opportunities for less invasive treatment options. This thesis explores whether the mitochondria and its related genes in the nuclear genome can be used as novel methods for the diagnosis and prognosis of cancers.

Aims and Methods:

Paper I: To investigate if mitochondrial dysfunction (characterized by mtDNA copy number variations) is associated with prevalent, incident cancer and cancer mortality – droplet digital PCR (ddPCR).

Paper II: To investigate the potential causal relationship between mitochondrial dysfunction (characterized by genetic predispositions in all mitochondrial-related genes) and common cancer risks – Mendelian randomization, colocalization.

Paper III: To investigate mitochondrial mutations as potential biomarkers for the early diagnosis of breast cancer – whole mitochondrial genome sequencing, bioinformatics, ddPCR.

Paper IV: To investigate the mitochondrial-related gene expression signature as a prognostic model to predict the clinical outcome for breast cancer patients – machine learning.

Results and conclusions:

Paper I: We found that mtDNA-CN was significantly associated with prevalent and incident cancer as well as cancer mortality. However, these associations were cancer-type specific and need further investigation.

Paper II: We identified potential causal relationships between mitochondrial-related genes and breast, prostate and lung cancer. Furthermore, this study identified candidate genes that can be the targets of potential pharmacological agents for cancer prevention.

Paper III: We comprehensively characterized the mtDNA mutation landscape of breast cancer biopsies and matched baseline whole blood samples. Notably, we have identified and validated mt.16093T>C mutation, which was associated with a 67% increased risk of developing breast cancer, and could potentially be used as early breast cancer diagnostic biomarkers.

Paper IV: We built a novel 14 genes mitochondrial signature model that could be an independent prognostic predictor and together with clinical variables as an improved model for predicting overall early-stage of breast cancer survival.

List of papers included in this thesis

Paper I

Li Y, Sundquist K, Wang X, Zhang N, Hedelius A, Sundquist J, Memon AA. Association of Mitochondrial DNA Copy Number and Telomere Length with Prevalent and Incident Cancer and Cancer Mortality in Women: A Prospective Swedish Population-Based Study. *Cancers (Basel). 2021 Jul 30;13(15):3842.*

Paper II

Li Y, Sundquist K, Zhang N, Wang X, Sundquist J, Memon AA. Mitochondrial related genome-wide Mendelian randomization identifies putatively causal genes for multiple cancer types. *eBiomedicine. 2023 Feb;88:104432.*

Paper III

Li Y, Sundquist K, Vats S, Hong M, Wang X, Chen Y, Hedelius A, Saal HL, Sundquist J, Memon AA. Mitochondrial heteroplasmic shifts reveal a positive selection of breast cancer. *Manuscript submitted*.

Paper IV

Li Y, Sundquist K, Wang X, Sundquist J, Memon AA. The mitochondrial signature for predicting outcome of early-stage breast cancer by machine learning. *Manuscript submitted.*

List of papers not included in this thesis

Vats S, Sundquist K, **Li Y**, Wang X, Hong MG, Sundquist J, Zarrouk M, Gottsäter A, Memon AA. Characterization of the Mitochondrial Genetic Landscape in Abdominal Aortic Aneurysm. *J Am Heart Assoc. 2023 Apr 18;12(8):e029248.*

Memon AA, Vats S, Sundquist J, **Li Y**, Sundquist K. Mitochondrial DNA Copy Number: Linking Diabetes and Cancer. *Antioxid Redox Signal. 2022 Dec;37(16-18):1168-1190.*

Popular summary

Cancer, a word that strikes fear into people, threatens millions of lives worldwide. What if we could detect cancer before it takes hold? We would be able to receive more effective treatments and improved our chances of survival. This is where the role of mitochondria, the tiny ancient bacteria that hacked our cells and now work as a powerhouse, comes into play.

In the quest to conquer cancer, early detection is key. This thesis delves into a groundbreaking exploration of whether mitochondria and their related genes could become game-changers in the battle against cancer. The journey begins with a crucial question: Can mitochondria and their genetic companions serve as novel tools for diagnosing and predicting cancer outcomes? With a focus on early detection, the thesis takes us on a scientific adventure through a series of studies.

The first part investigates mitochondrial dysfunction – a state where these cellular powerhouses are not functioning optimally. DNA that mitochondria house is known as mtDNA. By studying the changes in mtDNA copy number in healthy individuals and people with cancer, we uncover intriguing connections between mitochondrial dysfunction and cancer. These connections, however, were unique to different types of cancer, thus suggesting the complex nature of this relationship.

The second part delves even deeper, exploring the genetic codes that govern our mitochondria. Using a unique technique called Mendelian randomization based on the differences in our genetics that are dispersed by mother nature, we dive into the genes related to mitochondria and their potential impact on common cancers. In doing so, we unveil potential causal relationships and identify genes that could potentially drive cancer and become targets for innovative cancer prevention approaches.

In the third part, the focus shifts to the DNA mutations within mitochondria. These mutations are like tiny signposts, which might offer early clues to the presence of breast cancer. By comparing the mutations of mtDNA of breast cancer biopsies and cancer-free blood samples from the same patient, we uncover exciting clues. One specific mutation, known as mt.16093T>C, emerges as a potential biomarker for early breast cancer diagnosis. This discovery opens the door to more accurate and timely cancer detection.

Finally, the fourth part takes us into the realm of machine learning, where computers can discern hidden patterns within complex data. Here, we develop a model using 14 genes related to mitochondria. This model alone, or combined with clinical information, emerges as a powerful predictor of breast cancer outcomes in its early stages.

In summary, this thesis brings us to the frontier of cancer research, where mitochondria play a pivotal role in our battle against cancer. These findings offer hope for a future where early cancer detection becomes a reality, potentially leading to better treatment options and higher survival rates. Mitochondria – those small, hard-working, yet mighty, structures that might just hold the key to revolutionizing our approach to cancer detection and prognosis. The journey of science continues, fueled by the quest for a healthier, cancer-free world.

Abbreviations

MtDNA-CN	Mitochondrial DNA copy number
CI	Confidence interval
HR	Hazard ratio
OR	Odds ratio
ROS	Reactive oxygen species
DdPCR	Droplet digital polymerase chain reaction
NGS	Next-generation sequencing
COPD	Chronic obstructive pulmonary disease
CVD	Cardiovascular disease
SD	Standard deviation
ICD	International Classification of Diseases
OXPHOS	Oxidative phosphorylation
GWAS	Genome-wide association study
MR	Mendelian randomization
SMR	Summary-data-based MR
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
ER	Estrogen receptor
PR	Progesterone receptor
HER2	Human epidermal growth factor receptor
IVW	Inverse variance weighting
MR-PRESSO	MR Pleiotropy Residual Sum and Outlier
cfDNA	Cell-free DNA

HL	Heteroplasmy level
InDels	Insertions/deletions
AIC	Akaike information criterion
ROC	The receiver operating characteristic curve
AUC	The area under the ROC curve
DCA	Decision curve analysis
FPKM	Fragments per kilobase of exon model per million mapped fragments
TPM	Transcripts per kilobase of exon model per million mapped reads
GEO	Gene Expression Omnibus
Lasso	Least absolute shrinkage and selection operator
SCANB	Sweden Cancerome Analysis Network - Breast
TCGA-BRCA	The Cancer Genome Atlas Breast Invasive Carcinoma

Introduction

Cancer

Cancer is a leading cause of death and a complicated disease that can affect any part of the body. It begins when cells divide and spread in an abnormal and uncontrolled way. Based on where the cancer begins, cancer can be divided into five main types: carcinomas, brain tumors, sarcomas, leukemias and lymphomas. Although there are other cancer types, carcinomas that are capable of forming solid tumors stand out as the most common form of cancer.

At the level of cellular phenotype, different cancer types shared 14 commonalities in the process of forming malignant tumors. These include maintaining proliferative signaling, evading growth suppressors, undergoing non-mutational epigenetic reprogramming, escaping immune destruction, enabling replicative immortality, promoting tumor-related inflammation, interacting with polymorphic microbiomes, activating invasion and metastasis, inducing/accessing vasculature, inducing senescence in cells of different origins, causing genome instability and mutation, resisting cell death, reprogramming cellular metabolism, and unlocking phenotypic plasticity [1].

Current clinical methods primarily detect cancer when it becomes visible on imaging techniques, and a tissue biopsy remains the gold standard for diagnosis. Nonetheless, obtaining tissue samples necessitates invasive methods, which can lead to side effects such as pain, infection risk, bleeding, and extended recovery periods. Furthermore, a tissue biopsy has limitations in detecting molecular residual disease and may not accurately represent all tumors within the body. Research indicates that molecular changes occur well before the cancer becomes clinically visible, which suggests that early detection through molecular diagnosis can enhance patient survival and outcomes.

In addition, despite the progress in neoadjuvant and adjuvant therapies including chemotherapy, radiotherapy, hormone therapy, immunotherapy and targeted therapy, there remains a need to refine risk stratification and tailor treatment approaches for patients with cancer. Various clinical, histopathological and genetic factors have been employed to evaluate patient prognosis; however, more accurate and personalized markers are needed for improving patient outcomes, guiding treatment decisions, optimizing cancer care and reducing the burden of cancer on individuals and society.

Epidemiology

According to data from 185 countries and 36 cancers from GLOBOCAN estimates in 2020, there were 19.3 million cancer incidences and 10 million deaths. Female breast cancer (11.7%) has the highest incidence in the overall population followed by lung cancer (11.4%), colorectal cancer (10.0%), prostate cancer (7.3%) and stomach cancer (5.6%). Lung cancer (18.0%) has the highest mortality rate followed by colorectal cancer (9.4%), liver cancer (8.3%), stomach cancer (7.7%) and female breast cancer (6.9%). To be specific according to sex, lung cancer is the leading diagnosed cancer (14.3%) and the leading cause of mortality (21.5%) in males, while breast cancer is the leading diagnosed cancer (24.5%) and the leading cause of mortality (15.5%) in females [2]. With the increasing incidence and mortality rate of cancers, the economic cost worldwide from 2020 to 2050 has been estimated to be $25.2 trillion. Among the cancers, tracheal, bronchus, and lung (TBL, 15.4%) cancer, colorectal cancer (10.9%), breast cancer (7.7%), liver cancer (6.5%) and leukemia (6.3%) are responsible for the highest economic costs [3].
According to the statistics from The National Board of Health and Welfare in Sweden, from 1970 to 2021, cancer cases increased while mortality decreased for both women and men (Figure 1). There is a decreased trend in new cancer cases in 2020, which may potentially be due to delayed cancer diagnosis during the Covid-19 pandemic.

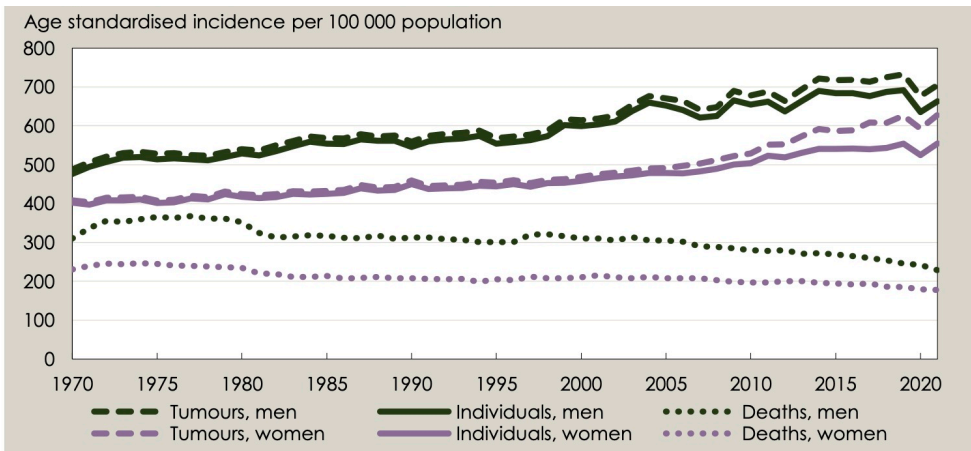


Figure 1. Number of new cancer cases and deaths in Sweden. Data source: The National Cancer Registry and the National Cause of Death Registry and The National Board of Health and Welfare. https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/statistik/2022-12-8309.pdf

As shown in Figure 2, in 2021, prostate cancer and breast cancer are the most common cancers for men and women, respectively, followed by skin cancer. Prostate cancer was the leading cause of death among men and 2,077 men died from

it, while lung cancer caused 1,871 deaths and was the leading cause of death among women.

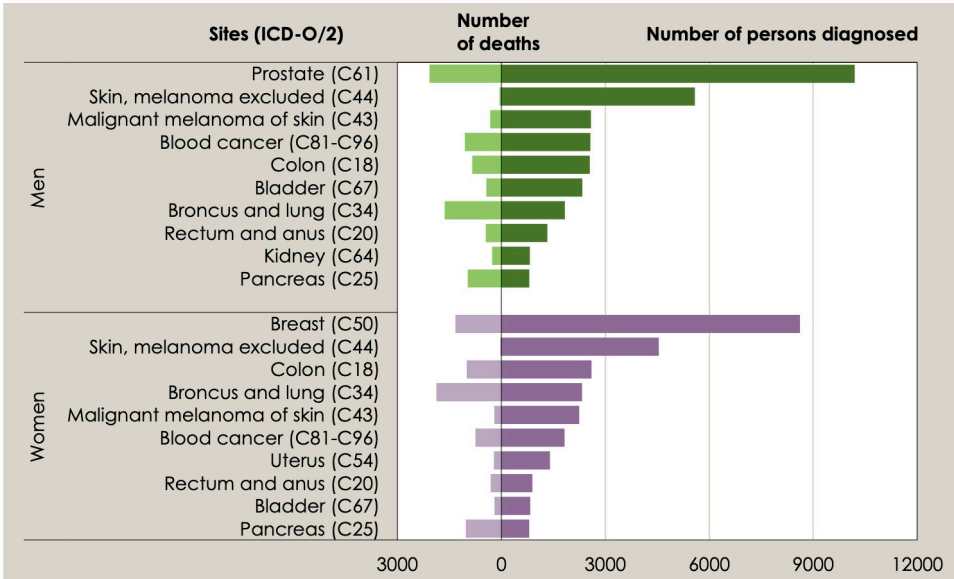


Figure 2. The ten most common cancer in Sweden, 2021. Data source: The National Cancer Registry and the National Cause of Death Registry and The National Board of Health and Welfare. https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/statistik/2022-12-8309.pdf

Risk factors

Cancer is normally caused by inherited genetic mutations (5% to 12%) and susceptible modifiable risk factors.

The inherited genetic mutations, which are known as germline mutations, were regarded as non-modifiable risk factors for certain cancers. For common cancers, patients with a first-degree family history of cancer have a 2- to 3-fold higher risk of developing the same cancer [4]. The advancements in genomic technologies such as genome-wide association studies (GWAS) and next-generation sequencing (NGS) facilitated the explosive identification and characterization of cancer susceptibility genes in population-based genetic studies. Over the past decade, cancer risk loci were widely identified in different cancers, such as the predisposition in *BRCA1, BRCA2, PALB2, TP53, ATM,* and *CHEK2* are associated with breast cancer; *BRCA1, BRCA2, RAD51C, RAD51D* are associated with ovarian cancer; *MLH1, MSH2, MYH, AXIN2, POLD1, POLE* with colorectal cancer; *BRCA2, HPC1, AR, VDR* with prostate cancer; *BRCA2, PALB2, CDKN2A* with pancreatic cancer; *BRCA2, CDKN2A, POLH* with melanoma; *CDH1* with gastric

cancer, et al [4-9]. The GWAS studies focusing on breast and prostate cancer have yielded the highest number of risk loci [10].
Cancer is a polygenic rather than single mutation caused disorder and all the identified cancer susceptibility loci to date are linked with moderate increases in risk, where odds ratios are generally below 1.5 [11]. There are more risk loci that are identified in different cancers and combining the effect of multiple risk loci by polygenic score may perform better for risk stratification [12].
Modifiable risk factors do not directly cause cancer and prevention by protecting from modifiable risk factors represents the most efficient strategy for managing the impact of cancers [13]. Here, the common modifiable risk factors for cancers as illustrated in Figure 3 and described below.

Figure 3. Major established modifiable cancer risk factors. Data source: The International Agency for Research on Cancer (IARC/WHO). https://cancerpreventioneurope.iarc.fr/preventable-cancers/

Tobacco consumption: stands as the foremost single cancer incident and mortality causing factor, which significantly contributes to the development of many cancers including but not limited to lung, oral, stomach, pancreas, cervix, bladder etc [14].
Alcohol consumption: accounted for 4% of cancer diagnoses worldwide and is associated with cancers of breast, colorectal, esophageal, liver, oral, pharynx and larynx in a dose-dependent manner [15,16].
Physical activity and diet: may contribute to obesity and hold significant influence as key determinants of cancer risk [17]. Sedentary behavior and high body fat are associated with major cancer-related outcomes. Diets rich in fruits and vegetables demonstrate protective effects against multiple cancer types, while excessive consumption of preserved meats has been associated with elevated cancer risk [18,19].
Infectious diseases: Certain factors causing infectious diseases contribute to cancer development. The bacterium *Helicobacter pylori* can cause peptic ulcers that then increases the risk of gastric cancer. Hepatitis B and C viruses (HBV, HCV) are

associated with the risk of liver cancer and human papillomavirus (HPV) infection is a leading cause of cervical cancer [20,21].
Exposure to carcinogens: exposure to ionizing radiation affects the immune system and paves the way for carcinogenesis, and exposure to ultraviolet radiation increases the risk of various skin cancers [22-24].
Gaining insights into the interplay between inherited genetic mutations, lifestyle and environmental factors is also crucial for tailoring personalized preventive strategies and interventions.

Diagnosis and prognosis

A timely cancer diagnosis can significantly enhance the prospects of effective therapeutic intervention. In Sweden, there are three national screening programs for the early detection of breast, cervical and colorectal cancer implemented by the National Board of Health and Welfare. The participant rate reached 85% for breast cancer screening and the coverage rate of cervical cancer was around 82-83% in 2021; the colorectal cancer screening program is under coordination. Patients who experienced cancer symptoms or suspicious results of a screening test need further diagnostic testing using approaches such as physical exams, laboratory tests, imaging-based tests and biopsy. For most cancers, a biopsy is the golden standard for diagnosis.
Although the benefits of early detection are widely acknowledged for the improvement of cancer outcomes across diverse populations, approximately 70% of cancer deaths are concentrated within low- and middle-income countries with a late diagnosis according to the statistics from the World Health Organization (WHO). The utilization of currently existing cancer diagnostic techniques including positron emission tomography (PET), computed tomography (CT), magnetic resonance imaging (MRI) and magnetic resonance spectroscopy (MRS), alongside molecular diagnostic techniques, has facilitated enhanced rates of early cancer detection. However, there are challenges facing early detection due to: the inadequate understanding of the intricate biology underlying cancer, the accurately determining risk for cancer development, the identification and validation of reliable biomarkers, the development of technologies sensitive enough for early detection and the evaluation of early detection methodologies [25]. Enhancing cancer diagnostics can be achieved by fostering multidisciplinary collaboration among radiologists, pathologists, genetic specialists and other specialties.
When diagnosed with cancer, the term prognosis estimates how the disease progresses. Many factors affect prognosis including age, sex, BMI, the type and origin of the cancer, subtypes, cancer stage and grade, and response to treatment. The TNM staging system, based on the integration of tumor size or depth (T), lymph node spread (N) and metastases state (M), is a foundation for the prognosis of cancer [26]. Cancer prognosis is assessed by the combination of clinical evaluation, diagnostic test, and predictive model. Gene expression, methylation, long non-

coding RNA, microRNA and protein were investigated to be potential diagnostic biomarkers to build predictive models for different cancers, peptide-based biomolecules and proteins are currently the most available cancer biomarkers. For breast cancer, the presence of estrogen receptors (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) was used to determine malignancy [27]. Plasma alpha fetoprotein (AFP) for the diagnosis of hepatocellular carcinoma [28], carcinoembryonic antigen (CEA) for lung cancer [29], prostate-specific antigen (PSA) for prostate cancer [30] etc.
With the advancement in cancer genetic research, nucleic acid based liquid biopsy approaches were developed for cancer detection and prognosis [31]. The liquid biopsy contains circulating tumor DNA (ctDNA), circulating tumor cells (CTCs), plasma cell-free DNA (cfDNA), extracellular vesicles, proteins and metabolites, that can be assessed as cancer biomarkers. The ctDNA, which included nuclear and mitochondrial DNA, accounts for 0.1-10% of the total cfDNAwith range from 10-100 ng/ml in plasma, and is widely investigated for clinical application [32,33]. Over the years, the Food and Drug Administration (FDA) has approved several liquid biopsy assays for the clinical application of cancer diagnosis including Epi proColon® for colorectal cancer diagnosis, cobas® EGFR Mutation Test v2 for non-small-cell lung cancer (NSCLC) diagnosis, therascreen PIK3CA RGQ PCR Kit for breast cancer diagnosis, the FoundationOne®Liquid CDx for the diagnosis of NSCLC, prostate, ovarian and breast cancers, and CellSearch™ test prognostic for breast, prostate, and colon cancer [34]. The liquid biopsy trials are continuously growing and around 200 trials started in 2022. Artificial intelligence (AI) methods were developed and applied to improve the performance of different liquid biopsy assays [35,36]. In the coming years, AI involvement in liquid biopsy analysis for cancer diagnosis and prognosis will ultimately improve patient outcomes and offer a brighter future in the fight against cancer.

Mitochondria

The mysterious mitochondria, as ancient bacteria invaded the eukaryotic cells and became key players during the long evolutionary history [37,38]. Mitochondrial is a double-membrane, bean-shaped structure, housing its mainly maternally inherited mtDNA containing 37 bioenergetic genes, but made of additional 1000s of nuclear genes. Besides generating adenosine triphosphate (ATP), the functions of mitochondria have been investigated extensively during the past decades.

Multifunction of mitochondria in cancer

Mitochondrial dysfunction has long been acknowledged as one of the cancer hallmarks. More than a century ago, Otto Warburg observed and demonstrated that

tumor cells tend to undergo glucose fermentation into lactate even in the presence of oxygen, rather than relying on oxidative phosphorylation for respiration and energy production [39]. Since then, the functions of mitochondria in cancer have been widely studied [40-44]. Apart from producing ATP to power life, mitochondria play crucial roles in multiple cellular processes including synthesizing metabolites, producing reactive oxygen species (ROS), maintaining redox balance and calcium homeostasis, regulating cell death and signaling, and controlling inflammation [45-48]. As shown in Figure 4, the multifaceted involvement of mitochondria in different stages of cancer pathophysiology makes it challenging to draw simple dogma about their roles [40].

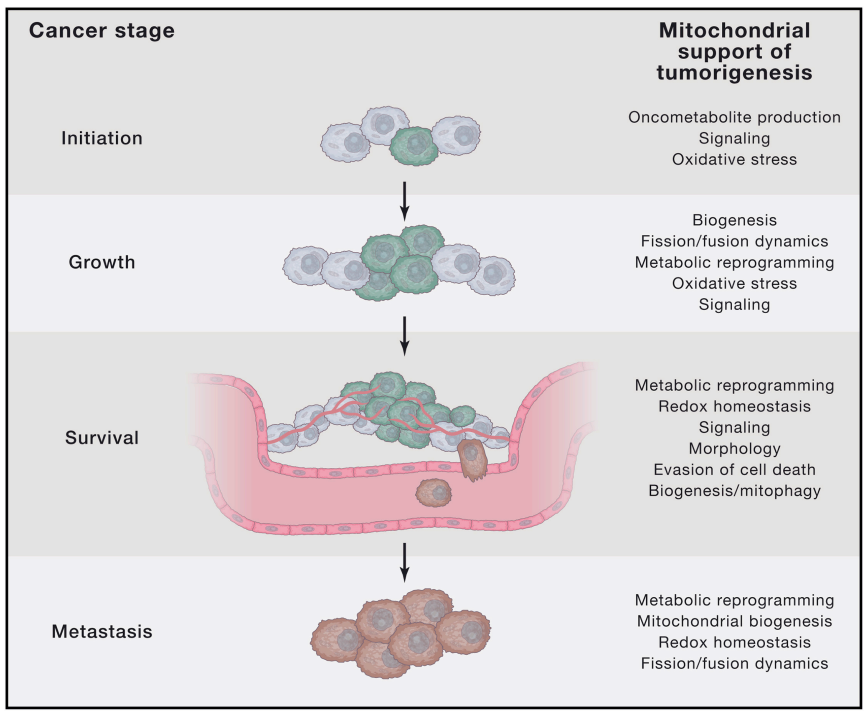


Figure 4. Mitochondria and stage of tumorigenesis. Reprinted from Cell. 2016 Jul 28;166(3):555-566. Vyas S, Zaganjor E, Haigis MC. Mitochondria and Cancer. © 2016 Elsevier Inc.

Mitochondria act as a stress sensor, allowing cells to modulate and adapt to the microenvironment. Tumor cells exploit this advantage to flexibly adjust and survive in harsh conditions such as hypoxia and nutrient deprivation, to promote tumorigenesis and cause chemoresistance [49,50]. However, the functions of mitochondria in different cancers vary due to the heterogeneity of tumor cells and microenvironments [51].

Metabolic reprogramming is a crucial process in cancer characterized by reciprocal feedback from mitochondrial metabolism. Mitochondria contribute to this process

by altering cell metabolism and generating excessive ROS, leading to mutagenesis and activation of the oncogenic signaling pathway, then accumulating oncometabolites to control gene expression and protein stabilization, conferring resistance to cell death or senescence [52]. Metabolic reprogramming is also regulated by mitochondrial dynamics, which encompasses alterations in their morphologies and mitochondrial trafficking, the contents exchange during the fission and fusion process, and the selective removal of damaged mitochondria [53]. Mitochondria are normally characterized by fragmented and rounded structures in a rich-nutrient environment. However, to survive conditions caused by glucose deprivation, mitochondrial fusion and biogenesis are activated to increase oxidative phosphorylation and NAD+ production to support the high energy demands of rapidly proliferating cells, which may trigger tumor cell immortalization [54]. In certain cancers, early hyperproliferative features may be characterized by decreased mitochondrial import and pyruvate metabolism, suggesting a preference for aerobic glycolysis [55]. The balance between glycolysis and oxidative phosphorylation (OXPHOS) in cancer is intricately dependent on tumor stage and type [56]. To cope with the metabolic chaos, cells may generate high levels of ROS that damage mitochondria, which must be removed through the mitophagy process to maintain mitochondrial mass homeostasis. Failure to do so can lead to the alteration of retrograde signaling controlled by mitochondrial metabolites, inducing cellular senescence and inhibiting mitochondrial biogenesis [57].

Of note, cancer prognosis and chemotherapy resistance have been associated with increased mitochondrial gain and OXPHOS during treatment. Mitochondria transfer between cells, through cell fusion, nanotube, gap junction, or extracellular vesicles, has been considered a communication phenomenon. Cancer cells have cunningly exploited this process by hijacking the mitochondria from other non-malignant cells to gain metabolic advantages and enhance proliferation and malignancy. Moreover, this process may severely impair immune cells, thus leading to immunosuppression and chemoresistance [58-61].

The presence of multiple copies of mtDNA due to its ploypoid nature allows the coexistence of inherited and mutated in a heteroplasmy state, which is the mixture of mutated and wild-type mtDNA within one cell [62,63]. The mitochondrial function is maintained when the deleterious heteroplasmic mutations are below a certain threshold. However, the shift to surpass this threshold leads to compromised mitochondrial performance and is commonly observed in cancers as pathogenic mitochondrial DNA mutations [64]. The mtDNA with high heteroplasmic mutations potentially causes the intensive proliferation of malignant cells, further leading to oncogenic or metastatic metabolism shift [65,66]. Studies have reported heteroplasmy shift as a potential driver for cancer progression and treatment response [67]. Thus, in cancer research, not only the presence of mtDNA mutations but the heteroplasmy shift matters, and early detection of those mutations with heteroplasmy shift will improve the cancer outcomes.

Overall, mitochondria play critical roles in various aspects of oncogenesis. The perturbation of the oncogenic signaling pathways coupled with alterations in mitochondrial functions, may potentially contribute to cellular malignant transformation, carcinogenesis, tumor progression and metastasis [40,52].

Potential clinical application of mitochondria in cancer

The mtDNA ranges from 10 to 10,000 in an origin-specific manner according to energy demands, thus also has a great abundance in circulation, allowing for the identification of cancer-associated mtDNA copy numbers (mtDNA-CN) variations, mtDNA mutations and alterations of mitochondrial-related nuclear genome. Efforts have been made to investigate the potential clinical application of mitochondria in cancer, by quantifying the mtDNA copy number (mtDNA-CN) in blood samples, identifying mtDNA mutations that are associated with cancer and pharmaceutically targeting mitochondria for cancer treatment. Some examples are described below.

For prevalent cancer, mtDNA-CN levels were higher in lymphocytic leukemia, lung and pancreatic cancer biopsies than matched normal tissue, while lower mtDNA-CN was observed in the biopsies of kidney and myeloid cancer compared to normal counterparts [68]. A study showed that elevated mtDNA-CN was associated with an increased risk of lymphoma and breast cancer, while decreased risk of skeleton and hepatic carcinoma cancer [69,70]. A comprehensive meta-analysis study of 39 publications underscored the association between high mtDNA-CN and, increased risk of hematological cancer (50%), decreased risk of soft tissue sarcoma (88%), endometrial (83%) and bladder cancers (49%) [71]. Our Swedish population-based prospective study observed that breast cancer had higher mtDNA-CN compared to controls and lower mtDNA-CN was associated with a 16% decreased risk of genital organ cancer [72]. Furthermore, elevated mtDNA-CN in peripheral blood was shown to be associated with a poor cancer prognosis, and elevated mtDNA-CN in cancer biopsies was associated with a better outcome [73].

The accumulation of mtDNA mutations reflects mitochondrial fitness and the mtDNA mutation has been identified as a major contributor of driver mutations in cancer [74]. For instance, germline mtDNA mutations mt.3197T>C, mt.13708G>A, mt.10398A>G, mt.16093T>C and mt.16519T>C were associated with breast cancer [75-77] and mt.16519T>C and mt.5460G>A was associated with pancreatic cancer [78-80]. Furthermore, somatic mtDNA mutations and their accumulation have been observed across multiple cancers, influencing cancer progression and metastasis [68,81]. However, compared to the nuclear genome, the mitochondrial genome is not extensively investigated in cancers. Considering the importance of mitochondrial mutations in cancer pathology, further investigations are needed to fully understand its role in cancer diagnosis and prognosis.

The majority of the genes governing the mitochondrial function and dynamics are synthesized in the nuclear genome and subsequently translocated into mitochondria [82]. In recent times, mitochondrial-related genes have been demonstrated as the

potential molecular biomarkers for colon cancer [83,84], renal cell carcinoma [85,86], prostate cancer [87], esophageal cancer [88], bladder cancer [89], lung cancer [90-93], liver cancer [94,95] and stomach cancer [96].
Although the mtDNA-CN, mtDNA mutation and mitochondrial-related genome alterations are not universal across all cancers, the abundance quantities of mitochondria in circulation presents a non-invasive alternative for cancer diagnosis and prognosis. Given the indispensable and multifaceted roles of mitochondria, numerous unresolved features remain. Rapid advancements in research and clinical trials have yielded promising results, offering stimulating prospects for the clinical application of mitochondrial medicine in the foreseeable future.

Advancements in cancer genetic research

In the past decade, the cancer genetic landscape research has been profoundly shaped by the emergence of cutting-edge technological and analytical methodologies. The identification and validation of cancer-associated molecular biomarkers using high throughput methods such as next-generation sequencing (NGS) and droplet digital PCR (ddPCR), have revolutionally improved our capacity to dissect the cancer molecular alteration with high resolution. Concurrently, data-driven analytical methods such as Mendelian randomization (MR) and machine learning, have enabled a deep exploration of complex causal associations and predictive molecular patterns in cancer.

High throughput experimental methods

In this thesis, we applied both NGS and ddPCR to identify and validate the cancer-associated mtDNA-CN variations and mutations.

Next-generation sequencing

Next-generation sequencing has revolutionized genomics research with the high-throughput resolution for deciphering the complexity of intrinsic genetic information. As illustrated in Figure 5, the NGS workflow of DNA sequencing contains four steps: library preparation, cluster amplification, sequencing, alignment and data analysis. Library construction begins with the fragmentation of DNA, adding specialized adapters to both ends of the DNA and then amplifying using PCR-based methods. The adapters contain complementary sequences that enable the fragments to attach to the flow cell. Then, add the prepared library to the flow cell and load it onto a sequencing machine, where the DNA sequences are further amplified into clonal clusters. The process called sequencing by synthesis starts, with modified nucleotides containing fluorescent tags and reversible terminators binding to the DNA template. The fluorescent signal captured by the machine

indicates a specific nucleotide has been added and the terminator is removed for the attachment of the next base. The sequencing process generates millions of short reads that constitute raw sequencing data. Following data acquisition, bioinformatics, including quality control, base calling, sequence alignment, variant calling and annotation, are applied to transform raw reads into informative genetic insights.

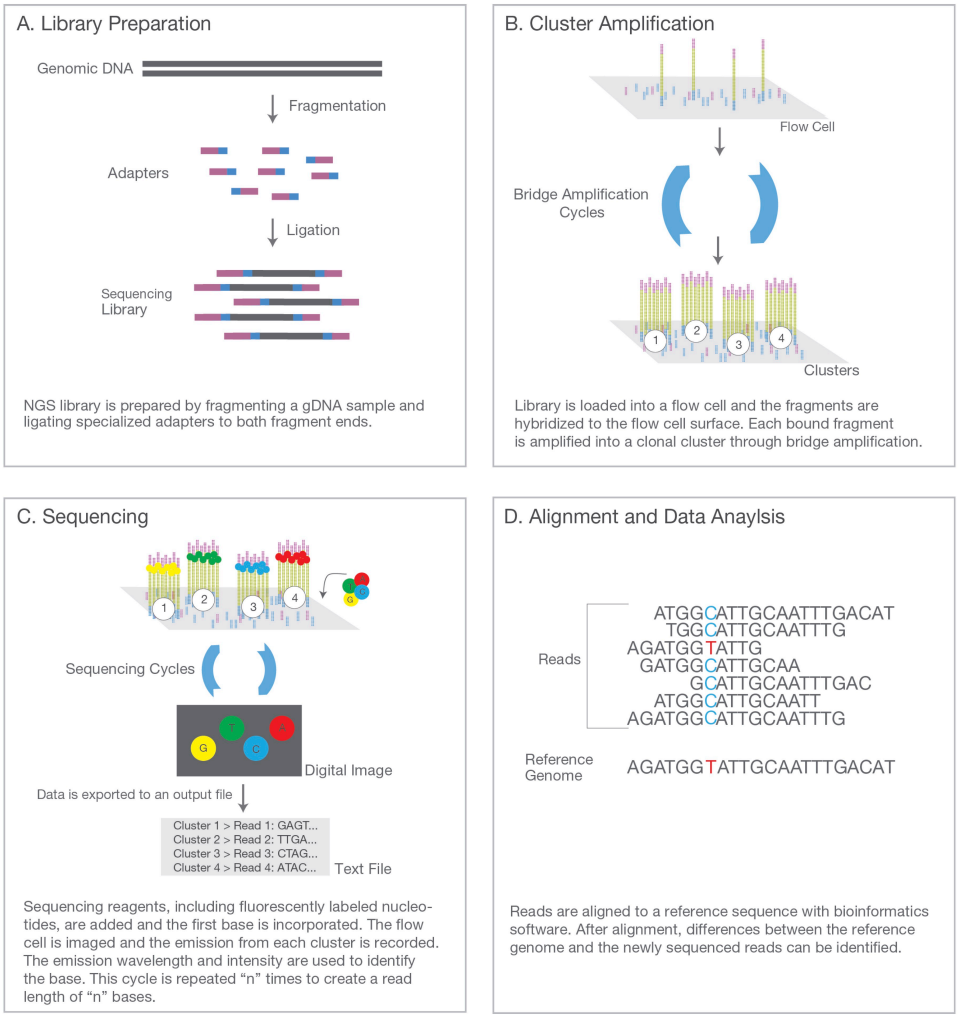


Figure 5. NGS workflow. Data source: Illumina, An introduction to Next-Generation Sequencing Technology.

Droplet digital PCR

NGS can only apply to a limited sample size with lower depth and is recommended to use for the identification of unknown molecular alteration. The ddPCR is cost-

effective when quantification of only a few targets is desired with high sensitivity. Compared to widely implemented read-time PCR, ddPCR provides highly accurate and direct quantification without requiring a calibration curve and quantitative variability, especially for low abundant targets [97,98]. Companies such as Bio-rad, Qiagan, Thermo Fisher Scientific, Stilla Technologies and RainDance Technologies, have provided ddPCR platforms. Here is the example of ddPCR workflow from Bio-rad, which starts with combining DNA samples, primers and probes with the ddPCR supermix, then loading the prepared samples into the disposable droplet generator cartridge. The cartridge is loaded to the droplet generator to make monodispersed droplets (~20,000). Subsequently, PCR amplification is performed that enables each nucleic acid molecule to be individually amplified with its exclusive droplet. After PCR, the plate can be loaded into the droplet reader, where the positive and negative droplets in each sample are plotted to the interface of the ddPCR software that allows visualization of each droplet and the target nucleic acid can be quantified as copies/µl (Figure 6).

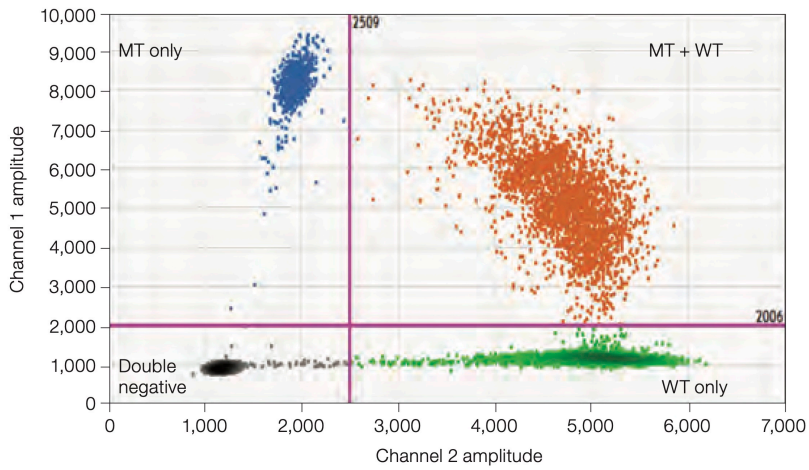


Figure 6. Classify droplets into clusters. WT: wild-type; MT: mutant. Data source: Bio-rad, Droplet Digital PCR Applications Guide.

Data-adaptive analytical methods

For large-scale genetic studies, MR estimates the causal relationships and machine learning for predicting outcomes.

Mendelian randomization

The golden standard to measure causal association is to perform a randomized control trial (RCT). Mendelian randomization (MR) was regarded as the RCT performed by mother nature, the genetic variants are intrinsically categorized that

can function as instrumental variables (IVs) to explore the potential causal association between exposure and outcome (Figure 7). In MR, using random allocation of alleles circumvents bias from unobserved modifiable variables such as lifestyle and environmental factors, as well as the intricate challenge of reverse causation [99].

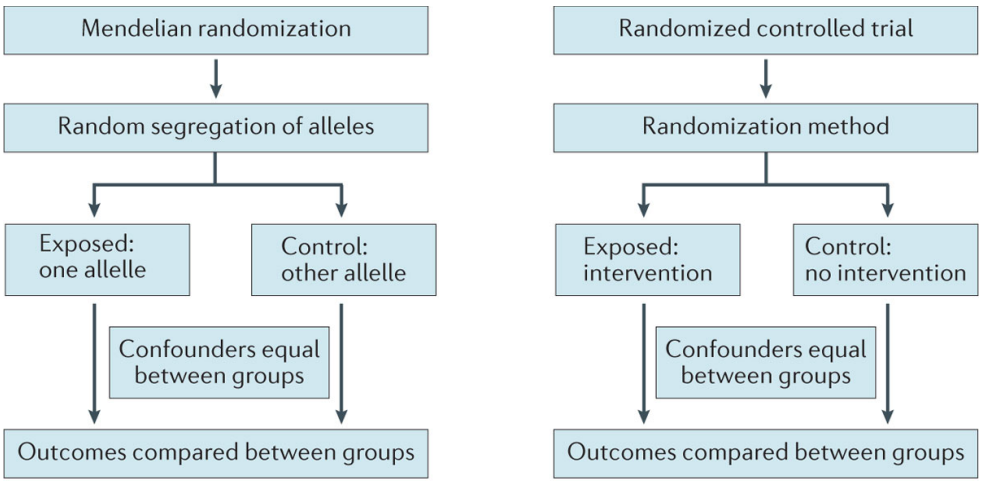


Figure 7. Comparison of the design of a Mendelian randomization study and a randomized controlled trial. Reprinted from Nat. Rev. Rheumatol. 2017 Feb 22;13(3):193. Robinson PC, Choi HK, Do R, Merriman TR. © The authors.

To perform MR analysis, adherence to three core assumptions is imperative as illustrated in Figure 8: 1. the genetic instruments must robustly associate with the exposure ($P < 5\times10^{-8}$); 2. the genetic instruments remain untethered to any confounding factors; 3. the genetic instruments impact on the outcome solely through exposure [99]. The two-sample MR is widely used that allows the evaluation of the association between IVs and exposure/outcome generated from distinct populations [100].

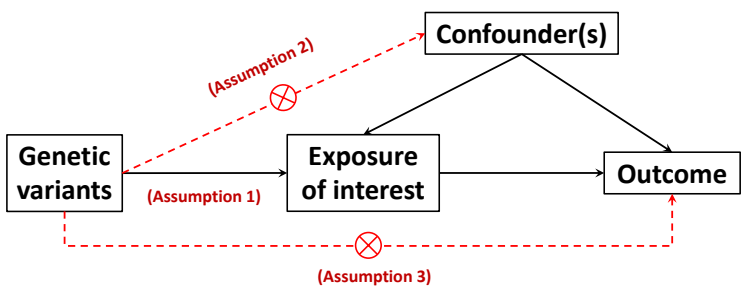


Figure 8. MR core assumption. Created with Powerpoint.

Machine Learning

Machine learning has emerged as a powerful and transformative tool in the healthcare and biomedicine field. It encompasses diverse algorithms that enable computers to learn intrinsic patterns and insights from complex and high-dimensional data, with the capacity to adapt and improve over time when new data is added in. Traditional statistic methods rely on making an inference about population-based samples, while machine learning employs the identification of patterns within data for repeatable predictions. Due to the heterogeneity and complex nature of cancer, machine learning has proven to be able to identify intricate relationships, unravel the complexities of cancer, and aid in diagnosis, prognosis, treatment decision and drug discovery.

Application and comparison of different machine learning methods are shown in Figure 9. In this thesis, we applied Lasso (Least Absolute Shrinkage and Selection Operator) Cox regression on the selected datasets to build a prognostic mitochondrial signature model for breast cancer. The Cox proportional hazard is a frequently used model to interpret coefficients in terms of the hazard ratio. However, the conventional Cox model is limited when we need to estimate the coefficient of numerous features. The Lasso regularization with a penalty term shrinks the coefficients associated with less important predictors towards zero, which can effectively promote variable selection and enhance the model's interpretability [101]. The Lasso Cox regression, which merges the principles of Cox regression analysis with the regularization technique, can effectively manage multicollinearity and can select relevant variables, mitigating overfitting, thus yielding an interpretable model for survival analysis.

Equation of Lasso Cox regression

$$f(x) = \lambda \sum_{j=1}^{p} |\beta_j| + \sum_{i=1}^{n} \left(d_i \log(\lambda) + \sum_{j=1}^{p} \beta_j x_{ij} \right) - \sum_{i=1}^{n} \log \left(\sum_{k \in R_i(t_i)} e^{\beta' x_k} \right)$$

n: the number of patients in the dataset.

p: the number of variables (in our study are genes).

λ: the regularization parameter that controls the amount of shrinkage applied to the coefficients.

β_j: the coefficient of the j-th predictor variable.

x_{ij}: the value of the j-th predictor for the i-th patient.

d_i: is an indicator variable that equals 1 if the death event occurs for the i-th patient.

t_i: the time at which the death event occurs for the i-th patient.

$R_i(t_i)$: the set of patients at risk at the time t_i.

Here, the first term in the equation ($\lambda \sum_{j=1}^{p} |\beta_j|$) is the Lasso penalty term. It shrinks the coefficients (β_j) towards zero; the second term, $\sum_{i=1}^{n}\left(d_i \log(\lambda) + \sum_{j=1}^{p} \beta_j x_{ij}\right)$, is the partial likelihood function of the Cox proportional hazards model that captures the relationships between predictor variables and the hazard function; the third term, $\sum_{i=1}^{n} \log\left(\sum_{k \in R_i(t_i)} e^{\beta' x_k}\right)$, is the log-likelihood term that accounts for the contribution of events that occurred for each patient, which quantifies the likelihood of observing the event times given the predictor variables and coefficients.

Lasso Cox regression aims to find the values of coefficients (β_j) that minimize the combined loss of the partial likelihood and the Lasso penalty, obtaining a model that balances predictive accuracy and the sparsity of predictors.

Method	Type of data	Example applications	Advantages	Disadvantages
Ridge (and LASSO/elastic) regression	Labelled Fixed number of features	Protein-variant effect prediction[122] Chemical/biochemical reaction kinetics[123]	Easy to interpret Easy to train Good benchmark	Cannot learn complex feature relationships Overfits with a large number of features
Support vector machine	Labelled Fixed number of features	Protein function prediction[124] Transmembrane-protein topology prediction[125]	Can perform both linear and non-linear classification and regression	Scaling to large datasets is often difficult
Random forest	Labelled Fixed number of features	Prediction of disease-associated genome mutations[126] Scoring of protein–ligand interactions[39]	Learns how important each feature is to the prediction Individual decision trees are human readable, allowing interpretation of how a decision is made Less sensitive to feature scaling and normalization so easier to train and tune	Less appropriate for regression Many decision trees are hard to interpret
Gradient boosting (for example, XGBoost)	Labelled Fixed number of features	Gene expression profiling[127]	Learns how important each feature is to the prediction Decision trees are human-readable, allowing interpretation of how a decision is made Less sensitive to feature scaling and normalization so easier to train and tune	Can struggle to learn underlying signal if noise is present Less appropriate for regression
Clustering	Unlabelled Fixed number of features	Differential gene expression analysis[15] Model selection in protein structure prediction[128]	For low-dimensional data, good clustering is easily identifiable Cluster validation metrics are available to assess performance	Scaling to large datasets is difficult for some methods Noisy datasets sometimes yield contradictory results
Dimensionality reduction	Unlabelled Large and fixed number of features	Single-cell transcriptomics[49] Analysis of molecular-dynamics trajectories[129]	Provides visual representation of data Goodness-of-fit evaluations usually available to assess performance	Hard to preserve both global and local differences in data Scaling to large numbers of samples is difficult for some methods
Multilayer perceptron	Labelled Fixed number of features	Protein secondary structure prediction[13] Drug toxicity prediction[54]	Can fit datasets with fewer layers than architectures such as convolutional neural networks, making it easier and faster to train	Easy to overfit Large number of parameters Hard to interpret
Convolutional neural network	Spatial data arranged in a grid; for example, 2D image (pixels) or 3D volumes (voxels) Allows variable input size	Protein residue–residue contact and distance prediction[23] Medical image recognition[24]	Variable input size Learns patterns irrespective of location in input	Receptive field, the amount of the input that is considered when predicting the output for each pixel, can be limited Hard to train deeper architectures that use many layers to increase the receptive field and make more complex predictions
Recurrent neural network	Sequential data (for example, biological sequences or time-series data) Allows variable input size	Protein engineering[68] Predicting clinical events[66]	Variable input size Sequences are found in many areas of biology	Long training times High computing memory requirements
Graph convolutional network	Data characterized by connections between entities (spatial, interaction or association) Allows variable input size	Predicting drug properties[77] Interpreting molecular structures[73,74] Knowledge extraction[130]	Variable graph sizes supported, which is important because most graphs in biology have variable size Learns patterns by following graph connectivity so predictor uses most relevant associations	High computing memory requirements for large, densely connected graphs Hard to train deeper architectures

Figure 9. Comparison of different machine learning methods. Reprinted from Nat Rev Mol Cell Biol. 2022 Jan;23(1):40-55. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. © 2021 Springer Nature Limited.

Aims

The overall aim of my thesis is to explore whether the mitochondria and its related genes in the nuclear genome can be used as novel methods for the diagnosis and prognosis of cancers.

The specific aims of each study were:

Paper I: To investigate if mitochondrial dysfunction (characterized by mtDNA copy number variations) is associated with prevalent, incident cancer and cancer mortality.

Paper II: To investigate the potential causal relationship between mitochondrial dysfunction (characterized by genetic predispositions in all mitochondrial-related genes) and common cancer risks.

Paper III: To investigate mitochondrial mutations as potential biomarkers for the early diagnosis of breast cancer.

Paper IV: To investigate the mitochondrial-related gene expression signature as a prognostic model to predict the clinical outcome of breast cancer patients.

Materials and methods

The objective of my research is to identify mitochondrial-associated biomarkers that can aid in the diagnosis and prognosis of cancer based on the Swedish (European) population, which involved clinical data analysis, experimental identification and validation using publicly available databases or cohorts.

Study population and data sources

WHILA cohort

Data used in paper I and paper III were derived from The Women's Health in Lund Area (WHILA) study, a prospective population-based cohort. WHILA was initiated in 1995 and targeted all women aged 50-59 years (born between 1935 and 1945) residing in the Scania region of southern Sweden. These women were invited to participate in a health survey following written informed consent and without financial compensation. Between December 1995 and February 2000, 6917 women out of the total population of 10,766, who lived in the five southern municipalities, underwent a physical examination. Participants were provided up to two hours to complete the questionnaire, which has been previously described [102], and in case of any uncertainties were encountered, experienced research nurses were available to offer guidance.

Participants were longitudinally followed from the day of inclusion until the occurrence of death or, no event occurred until 31 May 2015, whichever came first. The plasma samples were collected from all participants when they were recruited (baseline). However, the whole blood samples were collected midway through the study (starting from October 1997), a total of 3225 participants with the availability of their blood samples were included in the present cohort.

Paper I included all 3225 participants with blood samples, 3062 out of them were eligible for measuring the mtDNA copy number (mtDNA-CN) with ddPCR (Figure 10).

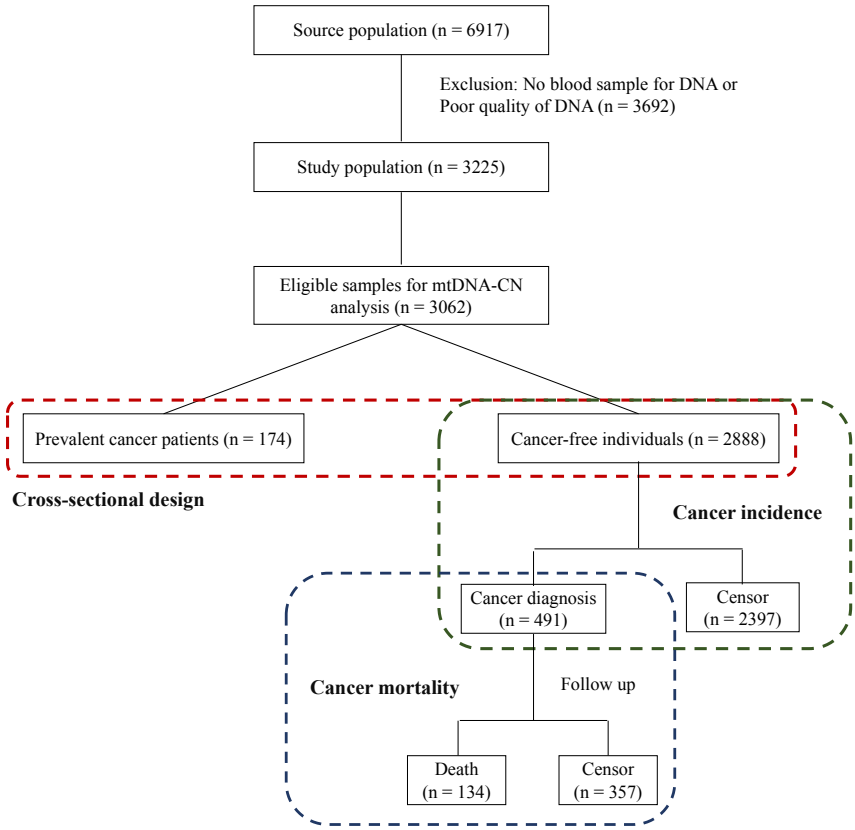


Figure 10. Flowchart of participants included in Paper I.

For paper III, samples were first selected for the mtDNA sequencing. We included a total of 349 women diagnosed with breast cancer during follow-up after excluding prevalent cancers. Among these women, 345 had baseline cancer-free plasma samples collected and 173 women had matched diagnostic biopsies. We included only those women who were diagnosed with breast cancer within three years of inclusion, as the detection of mutations before this period was not anticipated. Based on these criteria, 86 tumor biopsies were identified and where available, matched 50 baseline whole blood samples were included for mtDNA sequencing. After sequencing, we identified candidate mtDNA mutations and validated the mutations using ddPCR on the plasma samples collected from the same patients but before the clinical diagnosis of cancer. A total of 304 patients with available plasma samples were diagnosed with breast cancer during follow-up and 359 age and date-of-sampling matched controls were included for ddPCR validation (Figure 11).

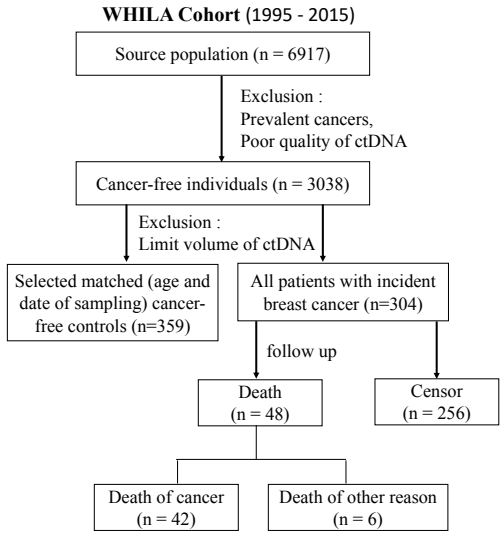


Figure 11. Flowchart of participants included in Paper III.

eQTL, mQTL and pQTL

In paper II, the eQTL summary statistics were retrieved from the eQTLGen Consortium, which was established to investigate the genetic architecture of blood gene expression and to understand the genetic factors underlying complex traits. The eQTLGEN Consortium included blood samples from 31,684 individuals and contains information on 10,317 trait-associated single nucleotide polymorphisms (SNPs) related to 19,960 genes [103]. The mQTL summary statistics containing information on SNPs that are robustly associated with gene methylation were downloaded from a meta-analysis study involving two cohorts (n = 1980) [104]. The pQTL summary statistics containing information on SNPs associated with the protein expression were downloaded from a meta-analysis study involving five proteome datasets [105-109]. However, all QTL datasets included in this study did not contain information on SNPs that are associated with the expression or methylation of genes located on the X and Y chromosomes and mtDNA.

GWAS summary statistics

In paper II, GWAS summary statistics of European ancestry investigating the association between SNPs and the risk of 18 types of cancers were obtained from publicly available large-scale studies.
The Breast Cancer Association Consortium (BCAC) comprised 133,384 breast cancer cases and 113,789 controls from 82 studies. The SNPs associations with breast cancer risk were adjusted for age, country and array-specific first 10 ancestry

principal components (PCs). In addition, we included the associations of SNPs with five molecular subtypes of breast cancer based on tumor grade and receptor status for further subgroup analysis [110].
For bladder cancer (cases: 2670, controls: 385,138), corpus uteri cancer (cases: 1515, controls: 210,164), esophagus cancer (cases: 992, controls: 386,907) and non-Hodgkin lymphoma (cases: 1579, controls: 386,126), SNP association were obtained from a published study of 454,787 UK Biobank participants. The SNPs association were adjusted for age, sex, array-specific 10 ancestry PCs, exome sequencing batch and sequencing-specific 20 PCs [111].
The SNPs associated with cervical cancer (cases: 1889, controls: 461,044) and melanoma (cases: 3598 controls: 459,335) risk were downloaded from the IEU OpenGWAS project derived from UK Biobank and adjusted for age, sex and genotype measurement batch [112].
For colorectal cancer (cases: 6581, controls: 463,421), gastric cancer (cases: 1029, controls: 475,087), pancreatic cancer (cases: 1196, controls: 475,049) and thyroid cancer (cases: 1054, controls: 490,920), SNPs associations were retrieved from a large cross-population study involving BioBank Japan, UK Biobank and FinnGen. The SNPs associations were adjusted for age, sex, top 20 genotype PCs and time from sampling to biobanking [113].
The SNPs associated with endometrial cancer (cases: 12,906, controls: 108, 979) risk were derived from a meta-analysis of 17 studies comprising of the UK Biobank, the Epidemiology of Endometrial Cancer Consortium (E2C2) and the Endometrial Cancer Association Consortium (ECAC), with adjustments for corresponding PCs in each study [114].
For kidney cancer, SNPs associations were obtained from a sex-specific GWAS study analysis of renal cell carcinoma for men (cases: 3227, controls: 4916) and women (cases: 1992, controls: 3095), with age, study center and significant eigenvectors as covariates [115].
For liver cancer, SNPs associations were obtained from a case-control GWAS study of alcohol-related hepatocellular carcinoma (cases: 775, controls: 1332), with adjustments for the first 10 genotype PCs, age, sex, and liver fibrosis [116].
For lung cancer, SNPs associations were obtained from a meta-analysis of 4 GWAS studies (cases: 11,348, controls: 15,861), involving individuals from the MD Anderson Cancer Center (MDACC), the Institute of Cancer Research (ICR), the National Cancer Institute (NCI) and the International Agency for Research on Cancer (IARC). The SNPs associations were adjusted for age, sex, histology and smoking status [117].
The SNPs associations of the oral cavity and pharyngeal cancer risk were obtained from a GWAS study (cases: 2497, controls: 2928) of the International Head and Neck Cancer Epidemiology Consortium (INHANCE), with adjustments for age, sex and PCs [118].
The SNPs associations of the overall ovarian cancer risk and different histotypes of epithelial ovarian cancer risk were obtained from the Ovarian Cancer Association

Consortium (OCAC, cases: 25,509, controls: 40,941) with the adjustments for project-specific PCs [119].
For prostate cancer, SNPs associations were obtained from a meta-analysis conducted by the Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome Consortium (PRACTICAL, cases: 79,148, controls: 61,106), with adjustments for PCs and study specific covariate [120].
GWAS summary statistics for SNPs associated with mtDNA-CN were obtained from a study based on UK biobank-scale whole exome sequencing of 415,422 individuals, with adjustments for age, sex, 40 ancestry PCs and exome sequencing batch [121].

SCANB and TCGA-BRCA

In paper IV, the breast cancer RNA-seq data were collected from Sweden Cancerome Analysis Network-Breast Initiative (SCANB) and The Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA). The SCANB started in August 2010 in the southern region of Sweden and patients with primary breast cancer were recruited. After signing the written informed consent, tumor biopsies, blood samples and data from the national quality registry for breast cancer were collected and followed. Until December 2020, SCANB has included over 16,000 patients and sequenced more than 13,500 RNA-seq libraries. In our study, we could obtain the RNA-seq data of 6657 patients from SCANB dataset with a follow-up of 4094 days [122]. The TCGA-BRCA data is part of a large-scale genomics program started in the United States in 2006 that included clinical data and molecular characterization of more than 20,000 primary cancer samples from 33 cancer types, alongside the matched normal tissue counterparts [123]. For breast cancer, we could obtain the RNA-seq data with sufficient clinical information of 1086 patients from TCGA-BRCA with a follow-up of 8065 days. However, after 4094 days, only 28 patients were followed, and five patients were deceased by the end of the study. In this case, to control the follow-up time for both datasets, those 28 patients were filtered out and 1039 breast cancer patients from TCGA-BRCA were included in this study.

Ethical statements

For paper I and paper III, data were derived from the WHILA cohort. The specific studies were covered in ethical approvals from the regional ethical committee at Lund University (approval nos. 2011/494 and 2015/6). All participants provided written informed consent following a comprehensive explanation of the study's objectives. The research was conducted in accordance with the guidelines in the Helsinki Declaration. Additionally, the studies did not include any vulnerable individuals, dependent relationships, or animal research.

For paper II and paper IV, all data used in analyses were generated from previous studies, for each of which ethical approval and individual consent were already obtained by their respective initiators/investigators. The data are freely available online and were permitted for public use.

Assessment of outcomes

The outcome in paper I and paper III, are the cancer incidence and mortality from the WHILA cohort. This information was obtained from the Swedish Cancer Registry and Death Registry and information on prevalent cancer was obtained from self-reported questionnaires. A total of 187 women were diagnosed with prevalent cancer at baseline, while 3038 women were cancer-free when they were enrolled at baseline. The cancer-free women were followed from the day of initial screening until the occurrence of one of the following events: (1) diagnosis of cancer; (2) death (from any cause: overall mortality, from cancer: cancer mortality); (3) end of the study (31 May 2015). For paper I, the cancer outcomes were further categorized based on the World Health Organization's International Classification of Diseases (ICD, revision 10) as: (a) breast cancer; (b) digestive system cancer (including liver, pancreatic gastric cancer, small intestine, colon, rectum, and oral cancer); (c) respiratory system cancer (lung cancer); (d) genital organ cancer (including ovary, cervix, uterus, and corpus cancer); (e) urinary system cancer (including kidney and urethral cancer); (f) hematological cancer (including myeloma, leukemia, and non-Hodgkin's lymphoma); (g) nervous system cancer; (h) melanoma and other malignant neoplasms of the skin; (i) endocrine gland cancer (thyroid cancer). For paper III, we only included breast cancer incidence as an outcome and identified a total of 304 women with incident breast cancer during a median follow-up period of 18.3 years. Among these women, 48 died from any cause (overall mortality) and 42 deaths were due to cancer (cancer mortality).
The outcomes for paper II are the 18 cancer summary statistics. For paper IV, the overall survival of breast cancer, was obtained from publicly available databases.

Assessment of covariates

For paper I and paper III, information on potential confounding factors was obtained from the health survey, which included the following factors: age at screening, body mass index (BMI), education level (categorized as 1-9, 10-11, or 12 years of schooling), alcohol consumption habits (categorized as no consumption, <12 g/day, or ≥12 g/day), smoking habits (classified as non-smokers, past smokers with <1 pack year and stopped smoking at least one month prior to the study, and current

smokers), physical activity at home (categorized as low activity and high activity) and physical activity at work (categorized as low activity, moderate activity and high activity). Information on the prevalent diseases and first-degree family history of cancer (defined as cases where the mother or sister had any type of cancer) were collected from baseline self-reported questionnaires and incident diagnoses of diabetes, cardiovascular disease and hypertension were obtained through the Swedish health registers.

Droplet digital polymerase chain reaction (ddPCR)

Quantification of mtDNA copy number

For paper I, ddPCR (QX200 AutoDG Droplet Digital PCR System, Bio-Rad, USA) was used to quantify the absolute mtDNA-CN out of the total genomic DNA extracted from the whole blood using QiAamp96 DNA Blood kit (Qiagen, Inc., Hilden, Germany). We used specific primers designed in the group to target the mitochondrial MT-ND1 (assay ID: dHsaCPE5029120) gene and the nuclear EIF2C1 (assay ID: dHsaCP1000002) gene. All probes had an Iowa Black® FQ quencher attached. Probes linked to HEX fluorophore were used for nuclear DNA, while mtDNA probes were attached to FAM fluorophore. The primer and probes were procured from Bio-Rad (Hercules, CA, USA). The runs along with quality control steps were performed as previously described procedures [124]. In brief, we pooled a 20 µl multiplex reaction containing 1 ng of DNA from samples, primers, probes, ddPCR Supermix for probes and 5U/reaction of the restriction enzyme HindIII. The reaction plate was sealed and incubated at room temperature for 20 minutes for restriction enzyme digestion. Subsequently, the plate was loaded into an automated droplet generator to generate droplets and followed by end-point PCR. The plate was then kept overnight at 4°C to maximize droplet recovery. The plate was later read on the droplet reader, data were downloaded and analyzed using QuantaSoft™ Software to determine the numbers of positive and negative droplets in each sample and the fraction of positive droplets was fitted to a Poisson distribution to calculate the absolute copy number in units of copies/µl. The final mtDNA-CN is presented as a ratio between absolute mitochondrial and nuclear (as reference) DNA copy number.

Validation of mtDNA mutations

For paper III, ddPCR was used to validate the candidate mtDNA mutations in the cell-free DNA (cfDNA) extracted from the plasma samples. The plasma samples were collected from whole blood after 2000 g for 10 min centrifugation within 8 h

of sample collection and stored at −80°C. We added a synthetic DNA template (TATAA Universal DNA Spike 166 bp; TATAA Biocenter) as spiked-in in each thrawed plasma sample (1ml). We obtained cfDNA following the manufacturer's instructions for the QIAamp Circulating Nucleic Acid Kit (Qiagen, Inc., Hilden, Germany). The Spike 166-bp assay facilitates the amplification of a specific 69-base segment within the synthetic template and this synthetic DNA was quantified in all samples by ddPCR. Samples with an extraction efficiency of more than 90% were included in the following analysis.
For the samples included in the mutation validation, we designed the mutation assays for mt.1888G>A and mt.16093T>C and ordered the assays along with primers, probes and reagents from BioRad. The primer details for all assays can be found below:

Mt.1888G>A:
MIQE Context: 4ec16488bb305da9800c1ad1d37d68aa|seq1:140-262:+
CAAGGACTAACCCCTATACCTTCTGCATAATGAATTAACTAGAAATAACTTTG
CAAGGAGA[G/A]CCAAAGCTAAGACCCCCGAAACCAGACGAGCTACCTAAGA
ACAGCTAAAAGAGCACACCCG

Mt.16093T>C:
MIQE Context: 4f645a5f181d564bb6a9baaf6fe0b62e|seq1:140-262:+
TGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCG
CTATGTAT[T/C]TCGTACATTACTGCCAGCCACCATGAATATTGTACNGTACCAT
AAATACTTGACCACCTGT

We pooled 22 µl ddPCR reaction composed of 11 µl of ddPCR Supermix for probes, 4 µl of template ctDNA, 1 µl of FAM- and HEX-labelled probes each, and 1 µl of 5U/reaction HindIII Restriction enzyme. The reactions were prepared in a semi-skirted 96-well plate and sealed using an automated pierceable foil heat sealer. Subsequently, a droplet was generated using the AutoDG system and the plate was resealed with the heat sealer. The PCR amplification was carried out with the following parameters: initial enzyme activation at 95 °C for 10 minutes, followed by 40 cycles of denaturation at 94°C for 30 seconds, annealing/extension at 54°C for 1 minute, and a final enzyme deactivation step at 98°C for 10 minutes. After PCR, the plate was incubated at 4 °C overnight to maximize droplet recovery. The plate was read using a droplet reader, and the data were analyzed using QuantaSoft Software to determine the number of droplets containing positive (mutant) and negative (wild-type) fluorophores for each sample.

Sequencing library preparation

In paper III, we sequenced the mtDNA from breast biopsies and matched available whole blood samples. After surgery, the biopsy samples were immediately snap-frozen and stored at –80°C in the South Swedish Breast Cancer Group tumor bank.

DNA from biopsies was extracted using the QiaAmp nucleic acid kit (Qiagen, Germany) according to the manufacturer's instructions. The standard EDTA blood tubes were used to collect baseline whole blood samples and blood genomic DNA was extracted using QIAamp96 DNA Blood (Qiagen, Germany) according to the manufacturer's instructions. Then, we applied long-range PCR using two pairs of primers to amplify the entire mitochondrial genome from extracted biopsy and whole blood DNA. The primer sequences employed are as follows:

MTL_Fwd1: AAAGCACATACCAAGGCCAC
MTL_Rev1: TTGGCTCTCCTTGCAAAGTT
MTL_Fwd2: TATCCGCCATCCCATACATT
MTL_Rev2: AATGTTGAGCCGTAGATGCC

The library preparations were carried out using the Nextera DNA Flex library preparation kit (Illumina, USA), following the manufacturer's guidelines. The concentrations of the purified libraries were measured using the Qubit 4.0 - 1× ds DNA high sensitivity assay (Invitrogen, USA), and the size range of the libraries (500 bp-1000 bp) was verified using Experion electrophoresis (Bio-Rad, USA). The final library pool with 2% internal control (phiX) was diluted to 80-100 pM and then loaded into our in-house iSeq100 system for sequencing. In the end, we obtained dual-index, paired-end raw fastq files as result.

Bioinformatic and statistical analysis

Mendelian randomization

In paper II, Mendelian randomization was used to investigate causal inference between mitochondrial dysfunction and cancers (Figure 12). We included the publicly available summary statistics from 18 common cancers (2107 to 491,974 participants), gene expression, DNA methylation and protein expression quantitative trait loci (eQTL, mQTL and pQTL, respectively. 1000 to 31,684 participants) on individuals of European ancestry, were included. Genetic variants located within 1000 kb on either side of the 1136 mitochondrial-related genes coding sequence (*in cis*) and robustly associated ($P < 5\times 10^{-8}$) 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. The SMR applied a two-step least-squares (2SLS) approach to estimate the effect size of an exposure on an outcome. In this study, the Linux version 1.0.3 of SMR software with default options was used to first extract the SNPs with $P_{\text{snp-mitodys}} < 5\times 10^{-8}$. We obtained 662,968 SNPs that were associated with the expression of 1013 mitochondrial-related transcripts from cis-eQTL, 931,304 SNPs that corresponded to 2550 mitochondrial-related DNA methylation

CpG sites from cis-mQTL and 23 SNPs that were associated with 23 mitochondrial-related protein expressions from cis-pQTL. Subsequently, those SNPs with linkage disequilibrium (LD) r-squared > 0.90 or < 0.05 were excluded after SMR analyses. The causal associations were calculated as follows:

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

Here, $\beta_{\text{mitodys-cancer}}$ represents the estimated effect size of mitochondrial dysfunction on cancer, $\beta_{\text{SNP-cancer}}$ represents the estimated effect size of the SNP on cancer and $\beta_{\text{SNP-mitodys}}$ represents the estimated effect size of the SNP on mitochondrial dysfunction. To control the genome-wide type I error rate, we used Bonferroni correction to adjust the SMR *P*-value. *P*-value thresholds of 0.05 / the number of probes (1013) = 4.936×10^{-5}, 0.05 / 2550 = 1.961×10^{-5} and 0.05 / 23 ≈ 0.002 were set for the statistical significances of the association between mitochondrial-related RNA expression, mitochondrial-related DNA methylation, mitochondrial-related protein expression, and cancer outcomes, respectively. Additionally, the embed heterogeneity in dependent instruments (HEIDI) test in SMR software was run to test whether the observed association was due to the vertical pleiotropy than the LD with the causal variant. The European ancestry obtained from the 1000 Genomes Project Consortium was used as a reference to estimate the LD [125]. A *P*-value threshold < 0.01 after the HEIDI test indicated the association was potentially due to linkage rather than pleiotropy and should be discarded from the analysis [126].
Following the primary SMR analysis, MR Egger, weighted median, Inverse variance weighting (IVW), simple mode and weighted mode were performed (TwoSampleMR R package) to further confirm the causal associations. The Cochran Q statistic embedded in MR Egger and IVW method was used to test heterogeneity across individual causal effects, and a Cochran's Q test *P-value* of < 0.05 indicates the presence of heterogeneity [127]. MR Egger regression and MR-PRESSO (Pleiotropy Residual Sum and Outlier) were applied to examine the presence of horizontal pleiotropy. MR Egger regression evaluates whether the pleiotropic effects of all identified genetic variants are independent of their instrument strength (InSIDE assumption). An intercept close to zero and a *P-value* > 0.05 indicates no directional pleiotropies driving the results of MR analysis [128]. MR-PRESSO can identify and adjust for outliers reflecting horizontal pleiotropic biases, a Global test *P-value* > 0.05 indicates the absence of horizontal pleiotropic outliers [129]. To access the robustness of the results, the leave-one-out sensitivity tests were applied, and one genetic variant was removed at each time to confirm that the results were not driven by a single variant (the estimate of the rest variants > 0) [130]. For the causal effect of a single genetic variant on the outcome, the Wald ratio estimate was applied to obtain the causal estimate [131]. Additionally, the strength of the variants was calculated using the F-statistic, where a value > 10 is considered a strong MR instrument.

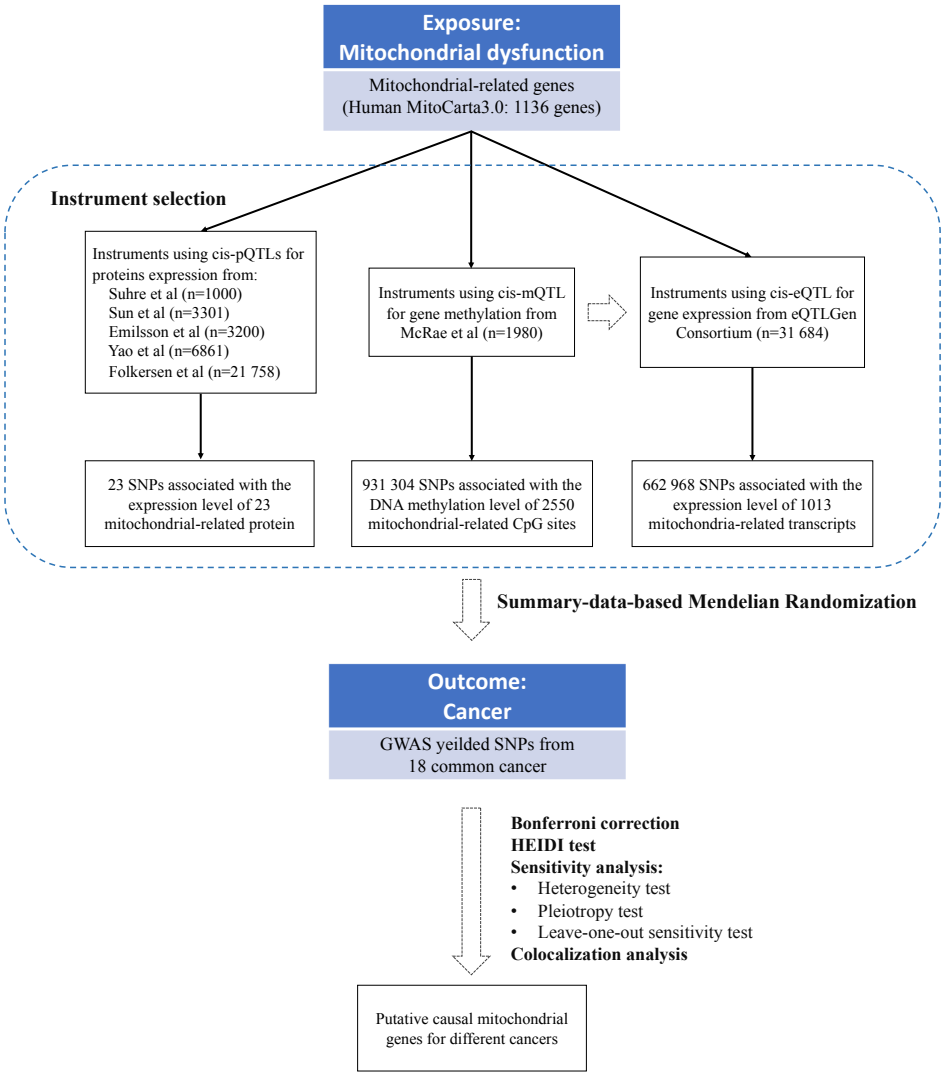


Figure 12. Workflow of Mendelian randomization analyses performed in Paper II

Importantly, to assess the presence of a shared causal variant in the region for both exposure and outcome, colocalization analysis was also conducted. The basic hypothesis for colocalization is:

H0: neither trait has a causal variant
H1: only trait 1 has a causal variant
H2: only trait 2 has a causal variant
H3: both traits have separate causal variants, but not the same variant
H4: both traits share the same causal variant

Except for the HEIDI test, we conducted another colocalization analysis based on the Bayesian test to estimate the posterior probability of shared variants [132]. In brief, for each leading SNP in the outcome GWAS summary statistics, we retrieved all SNPs within 100 kb up and downstream of the leading to analyze the posterior probability of H4 (PP.H4). We estimated the PP.H4 with default priors, assumed that the probability of a shared causal variant for trait 1 (P1) and trait 2 (P2) be 10^{-4}, and the probability of a shared causal genetic variant across both traits (P12) to be 5×10^{-5}. In this study, PP.H4 > 0.8 is the cut-off for the evidence of colocalization between the GWAS and QTL association.

MtDNA variant calling

In paper III, after sequencing of mtDNA, we obtained raw .fastq files and performed quality control on the files using MultiQC v1.8 [133], the adapters from library preparation were removed using TrimGalore v0.6.5 [134]. Then, the cleaned .fastq files were analyzed using the multifunctional integrated software MToolBox v1.0 [135]. This step generated a variant call format (.vcf) file containing detailed information on mtDNA mutations for each sample. The .vcf files were individually normalized and merged using BCFtools v1.8 [136] to create a final .vcf file that contained all samples. The effects of the mutations in the .vcf file were annotated by integrating the NCBI Homo sapiens Annotation with mitochondrial DNA function locations from MITOMAP [137]. Haplogrep v2.4.0 was utilized [138] to determine the haplogroup of each mtDNA sequence. Subsequent to variants identification, we performed quality control and defined heteroplasmic mutations (characterized by the ratio between mutant and wild-type alleles). The criteria of heteroplasmic mutation were adopted and modified according to Wei et al, 2019 [63], as follows:

1. Mutations with the lower bound of the confidence interval (CI) of heteroplasmic variant allele frequency (VAF) below 1% were retained.
2. Mutations at sites with multiple alternate alleles and alleles with a heteroplasmy level (HL) of less than 5% were discarded.
3. Heteroplasmic mutations at sites with over 98% of the upper bound of the CI were re-categorized as homoplasmic mutations.
4. Mutations falling within specific regions associated with misalignment errors related to homopolymeric tracts were discarded.
5. Heteroplasmic variants with sequencing depth below 100× or low HL (< 5%) with depth below 250× were discarded.

RNA sequencing data processing

In paper IV, the mRNA expression data of 6657 breast cancer patients in SCANB cohort in the format of fragments per kilobase of exon model per million mapped fragments (FPKM), with adjustment of library protocol, was downloaded from

Mendeley database (https://data.mendeley.com/datasets/yzxtxn4nmd). FPKM was further converted into transcripts per kilobase of exon model per million mapped reads (TPM) format then the values were normalized into log2 (TPM + 1). The mRNA expression data of 1039 breast cancer patients in TCGA-BRCA cohort in the format of log2 (TPM + 0.001) was downloaded and extracted from the XENA (https://xenabrowser.net/datapages/?dataset=TcgaTargetGtex_rsem_isoform_tpm&host=https%3A%2F%2Ftoil.xenahubs.net&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443), then normalized the values into log2 (TPM + 1).

Machine Learning

In paper IV, we aimed to construct mitochondrial-related gene expression signatures for breast cancer prognosis using machine learning. SCANB and TCGA-BRCA mRNA expression data and clinical information were employed to build the model. We first retrieved known 1136 mitochondrial-related gene list from MitoCarta v3.0 [139]. After excluding genes that were not presented in the SCANB mRNA datasets, a final set of 1094 genes was retained. We randomly divided the SCANB dataset into an 80% training set (n = 5326) and a 20% test set (n = 1331). Univariable Cox regression analysis was performed in the training cohort and 234 genes were identified to be significantly associated with the overall survival of breast cancer after Bonferroni correction ($P < 0.05/1094$). To ensure the robustness of the genes' selection, a bootstrap approach was performed by repeatedly sampling 80% of training sets 1000 times and 231 genes with $P < 0.01$ in over 800 times were retrieved. Then, we applied the Lasso Cox regression with 10-fold cross-validation (using the R package "glmnet") and shrunk to 33 genes in SCANB training dataset based on the "lambda.lse" value. Compared to the predictive coefficients of univariate Cox regression analyzed results on SCANB testing dataset and TCGA-BRCA dataset, a total of 14 genes that had the same predictive trends across all datasets were selected to build the most appropriate mitochondrial signature model. Based on the median expression level of the derived signature, we classified the patients into low and high-risk score groups for further analysis.

Statistics

For count/frequency comparison between two groups for categorical variables, Pearson chi-square test or Fisher's exact test was used depending on sample sizes; for continuous variables, student's t-test was used if the variables were normally distributed, otherwise, the Wilcoxon rank sum test (also known as Mann Whitney U Test) was used. A two-sided $P < 0.05$ was considered to reject the null hypothesis and achieve statistical significance in two group comparison. In paper I and III, for baseline characteristics comparison between prevalent cancer vs cancer-free or cancer-free vs incident cancer or mutation vs wild-type, Pearson chi-square test was

used to compare the categorical variables such as education level, smoking habits, alcohol consumption, activity at work, activity at home, diabetes, CVD, hypertension, first-degree family history of cancer. Student's t-test was used to compare the continuous variables such as age and BMI. In paper III, to compare the mtDNA mutations between biopsies and whole blood samples, Poisson regression was used to compare the difference between the mutation counts in two groups and the Wilcoxon rank sum test was used to compare the difference between mutation HLs in two groups. Fisher's exact test was performed to compare mutations between two groups across various mitochondrial regions or mutation consequences. The binomial test was used to assess the difference in the number of negative or positive mitochondrial HL changes within the same mitochondrial region. To account for multiple testing, Bonferroni correction was applied when conducting Fisher's exact test and the binomial test.

In paper I, to evaluate the association between prevalent cancer (yes/no) and mtDNA-CN at baseline, the linear regression analysis was performed and β coefficients were obtained. In paper II, we acquired β coefficients after SMR analyses and odds ratios (ORs) estimate per 1-ln increment in mitochondrial genome levels was used to measure the causal effect of mitochondrial dysfunction on cancer risk, which can be calculated by the formula: $OR_{mitodys\text{-}cancer} = \exp(\beta_{mitodys\text{-}cancer})$. In paper III, to evaluate the association between the presence of candidate mtDNA mutations (mt.1888G>A and mt.16093T>C) and breast cancer incidence, logistic regression analysis was performed, ORs and 95% confidence intervals (95% CIs) were calculated.

For the follow-up study design, Cox proportional hazards model was applied to explore the association between variable and event of interest by calculating the hazard ratios (HRs) with 95% CIs and potential confounders adjustment. In paper I, to evaluate the association between mtDNA-CN and cancer incidences, participants were dichotomized into high and low mtDNA-CN groups according to median value and the high group served as the reference group in the Cox regression analyses. The association between mtDNA-CN, all-cause mortality and cancer-specific mortality was also evaluated and competing risk models were applied while analyzing cancer-specific mortality to control the deaths from other causes. In paper III, the Cox proportional hazards model was performed and HRs with 95% CIs were calculated to evaluate the association between mutation mt.1888G>A, mt.16093T>C and cancer mortality. Kaplan-Meier survival curve allows the graphical presentation of survival probability over time that is commonly used in time-to-event studies. In paper IV, Kaplan-Meier analyses were performed to plot the overall survival according to the expression of mitochondrial signature in breast cancer patients. The time-dependent ROC curves with HRs and 95% CIs value of AUC (the area under the ROC curves) were plotted to evaluate the predictive performances of the mitochondrial signature models at 1-, 3-, and 5-year in different cohorts. To establish a nomogram model by integrating the mitochondrial signature with clinical features, stepwise regression analysis was performed to select the most

relevant variates from clinical characteristics including age, ER, PR, HER2, PAM50 subtypes and treatment. The model with the lowest Akaike information criterion (AIC) value was selected for nomogram establishment.

All statistical analyses were carried out in SPSS software version 23 (IBM, Armonk, NY, USA), SAS version 9.4 or R version 4.1.2 (https://cran.r-project.org/).

Results and discussion

Paper I

To investigate the association between mitochondrial dysfunction, characterized by mtDNA copy number (mtDNA-CN) variation, and prevalent, incident cancer and cancer mortality, we quantified the mtDNA-CN from baseline whole blood of all the eligible 3062 participants, 174 out of them had cancer before the inclusion. The cancer-free participants were followed with an average of 15.2 years untill the end of the study, 491 developed cancer and 134 died during the follow-up period (Figure 10).

To investigate the association between prevalent cancer and mtDNA-CN, we conducted linear regression analysis. As shown in Table 1, breast cancer was the most prevalent cancer (40.8%) in our study population. Our results demonstrated a significant positive association between prevalent breast cancer and elevated mtDNA-CN (12.39 copies/µl).

Table 1. The β coefficients and 95% confidence intervals of mtDNA-CN associated with prevalent cancer.

Characteristics	No. of Cases	Crude Univariate		Adjusted Univariate [a]	
		β (95% CI)	*P* Value	β (95% CI)	*P* Value
Prevalent cancer	174	2.75 (-2.67, 8.18)	0.320	4.18 (-1.18, 9.55)	0.059
Breast cancer	71	10.70 (2.36, 19.03)	0.012	12.39 (4.15, 20.63)	0.003
Digestive system	11	-6.53 (-27.52, 14.46)	0.351	-6.54 (-27.27, 14.20)	0.537
Respiratory system	1	3.86 (-65.64, 73.36)	0.913	1.27 (-67.36, 69.90)	0.971
Genital organs	30	-3.48 (-16.23, 9.27)	0.593	-0.48 (-13.09, 12.13)	0.941
Urinary system	5	-24.98 (-56.07, 6.11)	0.115	-20.77 (-51.47, 9.94)	0.185
Hematological cancer	7	-25.23 (-51.51, 1.05)	0.060	-25.21 (-51.19, 0.77)	0.057
Nervous system	5	-10.76 (-41.86, 20.34)	0.498	-8.80 (-39.53, 21.93)	0.575
Skin	17	11.27 (-5.62, 28.17)	0.191	9.51 (-7.19, 26.21)	0.264
Endocrine glands	17	-2.57 (-19.47, 14.47)	0.766	-0.88 (-17.58, 15.84)	0.918

[a] Adjusted for age, BMI, education level, smoking habits, alcohol consumption, activity at work and home, diabetes, CVD, hypertension.

For 491 participants who developed cancer, we conducted Cox regression analysis to investigate the association of the mtDNA-CN and cancer risk. Our findings, in Table 2 show that participants with lower baseline levels of mtDNA-CN had a 16% reduced risk (95% CI: 0.72, 0.98) of developing genital organ cancer (including

ovary, cervix, uterus, and corpus cancer) after adjustments for potential confounders. The baseline level of mtDNA-CN was also dichotomized into high (> 111 copies/μl, as reference) and low (≤ 111 copies/μl) groups according to the median value. Compared to the group with a high level of baseline mtDNA-CN, participants with lower mtDNA-CN had an increased risk of urinary system cancer (including kidney and urethral cancer with adjusted HR: 8.2, 95% CI: 1.06, 63.2) and hematological cancer (including myeloma, leukemia, and non-Hodgkin's lymphoma with adjusted HR: 1.97, 95% CI: 1.02, 3.81). However, no statistically significant association was found between mtDNA-CN and other cancer risks (Table 2).

Table 2. Hazard ratios and 95% confidence intervals of mtDNA-CN associated with incident cancer.

Characteristics	mtDNA-CN		HR (95% CI) per 1-SD decrease in mtDNA-CN
	High (*n* = 1406)	Low (*n* = 1482)	
All cancer			
No. of cancer diagnoses	242	249	
Person-years of follow-up	21,700	22,238	
IR, per 1000 person-years	11.15	11.19	
Crude HR (95% CI)	1 (Ref)	1.04 (0.87, 1.24)	1.01 (0.92, 1.10)
Adjusted HR (95% CI) [a]	1 (Ref)	0.99 (0.83, 1.19)	0.99 (0.90, 1.08)
Cancer types (adjusted HR and 95% CI [a])			
Breast cancer	1 (Ref)	1.05 (0.78, 1.42)	0.95 (0.84, 1.08)
Digestive system	1 (Ref)	0.83 (0.54, 1.29)	1.03 (0.84, 1.26)
Respiratory system	1 (Ref)	0.91 (0.42, 2.02)	1.18 (0.88, 1.60)
Genital organs	1 (Ref)	0.60 (0.33, 1.10)	0.84 (0.72, 0.98) [b]
Urinary system	1 (Ref)	8.20 (1.06, 63.2) [b]	1.08 (0.72, 1.63)
Hematological cancer	1 (Ref)	1.97 (1.02, 3.81) [b]	1.11 (0.85, 1.46)
Nervous system	1 (Ref)	1.62 (0.49, 5.31)	1.07 (0.66, 1.72)
Skin	1 (Ref)	0.82 (0.49, 1.39)	1.05 (0.82, 1.34)
Endocrine glands	1 (Ref)	0.79 (0.29, 2.13)	1.01 (0.63, 1.63)

[a] Adjusted for age, BMI, education level, smoking habits, alcohol consumption, activity at work and home, diabetes, CVD, hypertension.
[b] $P < 0.05$

For 134 participants who died during the follow-up, we conducted Cox regression analysis to investigate the association between the mtDNA-CN and cancer mortality. As shown in Table 3, we found a significant association between participants with lower baseline levels of mtDNA-CN had a 20% increased risk of all-cause mortality (95% CI: 1.01, 1.42) and similarly, a 21% increased risk of cancer-specific mortality (95% CI: 1.01, 1.45) after adjustment for potential confounders. Upon further stratification of the data based on cancer type, we

observed a notable negative association between the baseline level of mtDNA-CN and both all-cause mortality (adjusted HR: 2.15, 95% CI: 1.04, 4.44) and cancer-specific mortality (adjusted HR: 2.42, 95% CI: 1.03, 5.70) in patients with genital organ cancer. Similar associations were maintained when dichotomizing the mtDNA-CN level according to the median and compared to the high baseline mtDNA-CN group, low mtDNA-CN had significantly higher risks of mortality from all causes and cancer-specific in genital cancer patients. These results revealed a potential clinical application of mtDNA-CN as a prognostic marker for mortality of genital cancer patients.

Table 3. Hazard ratios and 95% confidence intervals of mtDNA-CN associated with mortality.

Characteristics	mtDNA-CN		HR (95% CI) per 1-SD decrease in mtDNA-CN
	High	Low	
All cancer			
No. of cancer patients	242	249	
No. of all-cause deaths	62	72	
No. of cancer-specific deaths	57	66	
Person-years of follow-up	1652	1578	
All-cause mortality rate, per 100 person-years	3.75	4.56	
Cancer specific mortality rate, per 100 person-years	3.45	4.18	
Adjusted all-cause mortality HR (95% CI)	1 (Ref)	1.14 (0.80, 1.62)	1.20 (1.01, 1.42) [b]
Adjusted cancer-specific mortality HR (95% CI) [a]	1 (Ref)	1.15 (0.80, 1.66)	1.21 (1.01, 1.45) [b]
Cancer types (adjusted all-cause mortality HR and 95% CI) [a]			
Breast cancer	1 (Ref)	1.42 (0.58, 3.51)	1.16 (0.77, 1.74)
Digestive system	1 (Ref)	1.08 (0.46, 2.51)	1.53 (1.02, 2.28)
Respiratory system	1 (Ref)	0.27 (0.06, 1.25)	0.48 (0.18, 1.24)
Genital organs	1 (Ref)	8.06 (1.75, 37.2) [b]	2.15 (1.04, 4.44) [b]
Urinary system	1 (Ref)	-	-
Hematological cancer	1 (Ref)	0.05 (0.00, 74.4)	-
Nervous system	1 (Ref)	-	-
Skin	1 (Ref)	-	-
Endocrine glands	1 (Ref)	-	-
Cancer types (adjusted cancer-specific mortality HR and 95% CI) [a]			
Breast cancer	1 (Ref)	1.71 (0.67, 4.35)	1.23 (0.86, 1.77)
Digestive system	1 (Ref)	0.89 (0.34, 2.39)	1.42 (0.90, 2.25)
Respiratory system	1 (Ref)	0.34 (0.08, 1.44)	0.48 (0.21, 1.08)
Genital organs	1 (Ref)	5.59 (1.61, 19.4) [b]	2.42 (1.03, 5.70) [b]
Urinary system	1 (Ref)	-	-
Hematological cancer	1 (Ref)	-	-
Nervous system	1 (Ref)	-	-
Skin	1 (Ref)	-	-
Endocrine glands	1 (Ref)	-	-

[a] Adjusted for age, BMI, education level, smoking habits, alcohol consumption, activity at work and home, diabetes, CVD, hypertension.
[b] $P < 0.05$

To our knowledge, this remains the first population-based prospective cohort study that investigated the association between mtDNA-CN and cancer prevalence, incidence and mortality in multiple cancer types. Compared to our well-optimized and accurate ddPCR method for the quantification of the absolute mtDNA-CN, real-time PCR and whole genome sequencing (WGS) are still the most commonly used methods to semiquantify the mtDNA-CN [140,141], even for the latest cancer studies [142-144]. Our study cohort only included middle-aged women that were followed prospectively for up to 20 years, which minimized confounding factors related to age and sex. However, the study's sample size was limited to specific cancer types with a small number of cases and affecting statistical power. The findings of this study indicate that mtDNA-CN is associated with prevalent and incident cancer and cancer mortality in a cancer-specific manner. After this publication, another group performed a large-scale meta-analysis that included more than 300,000 participants and, in agreement with our conclusions confirmed that the association of mtDNA-CN and cancer is highly heterogeneous and, in a cancer-specific manner [71]. We also published a review paper that summarized the associations between mtDNA-CN and cancers [140]. Nevertheless, no definitive conclusion could be drawn on the association between mtDNA-CN and cancer when considering all cancer types, and the possible explanation for the heterogenous and conflicting associations observed in different studies is the variability in sample selection, study design and methods employed to measure mtDNA-CN.

Considering the dynamic feature of mtDNA-CN caused by genetic predisposition and exposure to environmental and lifestyle factors [121,145], alternative approaches based on genetic variants could be adopted to uncover the association especially the causal correlation between mtDNA-CN and cancers. To address this, we later investigated the causal effects of mtDNA-CN variation on cancers using Mendelian randomization. The results showed that the causal association of mtDNA-CN on cancers was also cancer type specific and mtDNA-CN was causally associated with the risk of cervical and ovarian cancer (genital organ cancer) [77], which allied with the finding of this study.

Paper II

To identify the causal relationship between mitochondrial dysfunction characterized by genetic predisposition (SNPs) in all known mitochondrial-related genes and 18 common cancers of European ancestry, Mendelian randomization (MR) and colocalization analyses were performed (Figure 12).

We first conducted large-scale two-sample MR analyses on mitochondrial-related cis-eQTLs and cancer outcomes (Figure 13). Out of 18 cancers, we identified a total of 7 associations across 7 distinct SNPs for breast cancer, 4 associations across 4 distinct SNPs for prostate cancer, one association for gastric cancer and no

significant genetic association for the other cancer types. The colocalization analyses confirmed the causal effects of mitochondrial gene expression related SNPs on breast and prostate cancer. In breast cancer, one standard deviation (SD) decrement of *FDPS* expression caused by rs6677385 was associated with a 34% decreased cancer risk (95% CI: 0.49, 0.83); one SD increase of *NSUN4* expression caused by rs41293273 was associated with a 5% increased cancer risk (95% CI: 1.03, 1.07). In prostate cancer, one SD increase of *NSUN4* expression caused by the same SNPs in breast cancer was associated with a 6% increased cancer risk (95% CI: 1.03, 1.09).

Type of cancer	Gene	Instrumental variant	eQTL size effect[a]	OR (95% CI)	P_{SMR} value	Colocalization[b]
Breast cancer						
	FDPS	rs6677385	-0.09	0.66 (0.49 - 0.83)	9.77×10^{-7}	0.89
	HSCB	rs6519752	0.08	1.78 (1.52 - 2.05)	1.57×10^{-5}	0.74
	MPC2	rs203777	-0.03	1.39 (1.25 - 1.54)	1.08×10^{-5}	0.004
	MRPS18C	rs1565909	0.14	1.34 (1.25 - 1.43)	1.09×10^{-10}	< 0.001
	NSUN4	rs41293273	0.03	1.05 (1.03 - 1.07)	5.24×10^{-6}	0.90
	PARK7	rs35675666	-0.04	0.91 (0.87 - 0.95)	2.15×10^{-6}	0.003
	YBEY	rs62224180	0.78	1.04 (1.02 - 1.06)	3.05×10^{-5}	< 0.001
Prostate cancer						
	CASP8	rs7560328	-0.21	1.22 (1.15 - 1.30)	2.89×10^{-7}	< 0.001
	NSUN4	rs41293273	0.66	1.06 (1.03 - 1.09)	1.01×10^{-5}	0.94
	SLC25A37	rs2928682	-0.49	1.13 (1.10 - 1.17)	2.84×10^{-12}	< 0.001
	UQCC1	rs4911178	-0.46	1.09 (1.06 - 1.13)	8.70×10^{-7}	0.004
Gastric cancer						
	BAK1	rs210143	0.67	0.87 (0.81 - 0.92)	6.55×10^{-7}	< 0.001

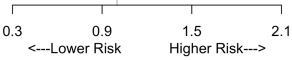


Figure 13. Mendelian randomization and colocalization analyses identified causal relationships between the expression of mitochondrial-related genes and breast, prostate cancer.

[a] Represents the effect size (β) of a variant on mRNA expressions. $\beta > 0$ means positive association, and $\beta < 0$ means negative association.

[b] 'Colocalization' indicates PP.H4 between eQTLs and cancer outcomes, PP.H4 > 0.8 was applied as the cut-off for the evidence of colocalization.

We then conducted MR analyses on mitochondrial-related cis-mQTLs and cancer outcomes (Figure 14). Out of 18 cancers, we identified a total of 15 associations across 14 distinct SNPs for breast cancer, 11 associations across 10 distinct SNPs for prostate cancer, one association for gastric cancer, 4 associations across 3 distinct SNPs for lung cancer and 2 associations in SNPs for melanoma. The colocalization analyses confirmed the causal effects of mitochondrial gene methylation related SNPs on breast, prostate and lung cancer. In breast cancer, we identified 6 distinct SNPs that can regulate *NSUN4* methylation levels at 7 different CpG sites and all exhibited a positive association with cancer risk. In prostate cancer, we found the same positive association between *NSUN4* methylation and cancer risk. Additionally, one SD increase of *NUDT5* methylation caused by rs4750175 was associated with a 4% decreased prostate cancer risk (95% CI: 0.95, 0.98). In lung cancer, one SD increase of *VARS2* methylation caused by SNPs

presented in different methylation CpG sites was associated with an over 20% increased cancer risk.

Type of cancer	Gene	Methylation probe	Instrumental variant	mQTL size effect[a]	OR (95% CI)	P_{SMR} value	Colocalization[b]
Breast cancer							
	BCL2L11	cg09907170	rs73954926	0.82	1.07 (1.04 - 1.11)	1.08×10^{-5}	0.003
		cg27608154	rs73954941	-1.51	0.96 (0.95 - 0.98)	2.56×10^{-6}	0.003
	MRPL23	cg07977153	rs4929956	0.26	0.85 (0.78 - 0.92)	1.32×10^{-7}	< 0.001
	NSUN4	cg14993813	rs6682266	-0.35	0.91 (0.87 - 0.96)	1.50×10^{-5}	0.96
		cg17875957	rs5013329	-0.40	0.92 (0.89 - 0.96)	9.86×10^{-6}	0.96
		cg17806798	rs41293277	-1.12	0.97 (0.96 - 0.99)	5.34×10^{-6}	0.96
		cg04241075	rs41293277	-1.04	0.97 (0.96 - 0.98)	5.59×10^{-6}	0.96
		cg06741803	rs56063031	-0.36	0.92 (0.88 - 0.96)	1.56×10^{-5}	0.96
		cg00530320	rs6681857	0.47	1.07 (1.04 - 1.10)	5.81×10^{-6}	0.96
		cg15580309	rs111226885	1.19	1.03 (1.02 - 1.04)	1.72×10^{-6}	0.96
	SLC25A22	cg23587532	rs7928917	0.23	0.85 (0.78 - 0.92)	6.55×10^{-6}	< 0.001
		cg11475788	rs61876744	0.21	0.83 (0.75 - 0.91)	8.67×10^{-6}	< 0.001
	SLC25A44	cg19263494	rs72708291	-0.84	0.96 (0.95 - 0.98)	6.58×10^{-6}	< 0.001
	TRMT1	cg22237401	rs74569397	0.66	1.12 (1.08 - 1.17)	1.40×10^{-6}	0.007
		cg12014333	rs16995252	0.59	1.14 (1.08 - 1.19)	3.22×10^{-6}	0.007
Prostate cancer							
	BIK	cg07972488	rs2294413	0.20	1.48 (1.31 - 1.65)	2.76×10^{-6}	< 0.001
	NSUN4	cg04241075	rs41293277	-1.04	0.96 (0.95 - 0.98)	1.16×10^{-5}	0.96
		cg17806798	rs41293277	-1.12	0.97 (0.95 - 0.98)	1.11×10^{-5}	0.96
		cg15580309	rs111226885	1.19	1.03 (1.02 - 1.05)	7.95×10^{-6}	0.96
	NUDT5	cg22687873	rs4750175	0.03	0.96 (0.95 - 0.98)	1.15×10^{-5}	0.99
	TSPO	cg13160331	rs138909	0.19	1.40 (1.26 - 1.54)	2.72×10^{-6}	< 0.001
	VARS2	cg02186769	rs2233959	0.24	0.84 (0.76 - 0.92)	1.63×10^{-5}	0.009
		cg05103231	rs2524108	0.24	0.80 (0.71 - 0.90)	1.24×10^{-5}	< 0.001
		cg12457901	rs2233980	0.86	0.93 (0.90 - 0.96)	2.19×10^{-6}	0.009
		cg16113650	rs2249935	0.20	0.70 (0.54 - 0.86)	1.24×10^{-5}	< 0.001
		cg26467571	rs9264490	0.28	0.83 (0.77 - 0.91)	2.34×10^{-6}	< 0.001
Gastric cancer							
	BAK1	cg00700324	rs511515	-0.28	1.40 (1.25 - 1.55)	1.41×10^{-5}	< 0.001
Lung cancer							
	VARS2	cg12457901	rs2523593	0.84	1.22 (1.14 - 1.30)	1.18×10^{-6}	0.98
		cg14935711	rs2523593	0.76	1.24 (1.15 - 1.33)	1.41×10^{-6}	0.98
		cg15848685	rs2596500	0.80	1.23 (1.14 - 1.31)	2.81×10^{-6}	0.98
		cg16958594	rs2596495	0.75	1.22 (1.13 - 1.31)	7.23×10^{-6}	0.98
Melanoma							
	SPG7	cg09560549	rs8060502	0.39	1.003 (1.002 - 1.004)	2.21×10^{-8}	< 0.001
		cg15206445	rs8060502	0.44	1.003 (1.002 - 1.004)	6.88×10^{-9}	< 0.001

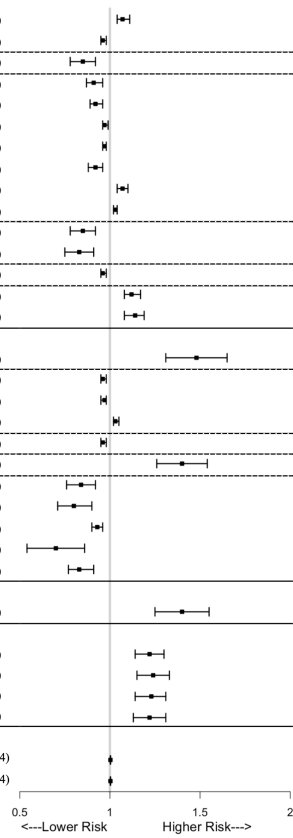


Figure 14. Mendelian randomization and colocalization analyses identified causal relationships between the methylation of mitochondrial-related genes and breast, prostate and lung cancer.

[a] Represents the effect size (β) of a variant on DNA methylation. $\beta > 0$ means positive association, and $\beta < 0$ means negative association.

[b] 'Colocalization' indicates PP.H4 between mQTLs and cancer outcomes. PP.H4 > 0.8 was applied as the cut-off for the evidence of colocalization.

We also conducted MR analyses on mitochondrial-related cis-pQTLs and cancer outcomes. However, no causal association was identified, which might be attributed to the incompleteness of the pQTL datasets, thus limiting the identification of SNPs robustly associated with mitochondrial-related protein expression.

The numerous roles of mitochondria have been demonstrated over a century, and the increasing numbers of studies have not surprisingly reported the relevance of mitochondrial dysfunction in different cancers ranging from mtDNA mutation, mtDNA-CN to molecular alteration in selected mitochondrial-related nuclear genes [80,146,147]. However, earlier studies did not investigate causal inferences between

mitochondrial dysfunction and cancers. This data-driven study fills the gaps by conducting Mendelian randomization and examines the causal associations of mitochondrial dysfunction characterized by genetic predisposition in common cancers. We identified specific mitochondrial-related genes and proved their causal relationship with cancer. Notably, the alteration of, *FDPS* was associated with breast cancer, *NUDT5* with prostate cancer, *VARS2* with lung cancer, and *NSUN4* with both breast and prostate cancers. However, except for *NSUN4,* the other three genes were relatively well-studied and regarded as anti-cancer drug targets for cancer prevention and treatment [148-150]. Untill now, the mechanistic studies on *NSUN4* have mainly been focused on its functions in the methylation process [151,152]. Here, we identified three commonly shared SNPs related not only to *NSUN4* expression but also to its methylation that were causally associated with both breast and prostate cancer after ruling out the horizontal pleiotropy, underscoring and affirming the important role of this gene involved in the carcinogenesis of hormone-dependant cancers. How the SNPs presented in *NSUN4* contribute to carcinogenesis warrants further investigation on additional cohorts and experimental studies.

Mitochondrial dysfunction, is a general concept but a complicated process, which is challenging to measure directly. The present study included all known mitochondrial-related genes from the nuclear genome, thereby avoiding selection bias and directly addressing the mitochondrial dysfunction caused by nuclear genetic predisposition. However, there is still no QTL dataset developed specifically containing the association of mtDNA genetic variants and mitochondrial gene expression, methylation or protein expression. Furthermore, there is no GWAS dataset that can directly address mitochondrial dysfunction, thus our assessment of mitochondrial dysfunction is only limited to nuclear genetic predisposition, hence potentially underpowered to detect the direction of causal association and unable to estimate the direct causal effects by performing multivariable MR. For the causal association analysis between the methylation of mitochondrial-related genes and cancer, we identified more association signals across distinct genetic loci even after colocalization analysis due to the mQTL dataset employed in this study providing all SNPs assigned to a gene. However, the eQTL dataset provided only the most significant SNP (topSNP) assigned to a gene, thus potentially filtering out possible causal genetic variants for cancer. Nevertheless, this is the first comprehensive MR study that included a large sample size (up to 491,974 participants) and demonstrated the causal effects of mitochondrial dysfunction on breast, prostate and lung cancer. Moreover, the identified causal genes could be potentially considered for inclusion in genetic screening for cancer prevention.

Paper III

To investigate if the mitochondrial mutations could potentially be biomarkers for early diagnosis of breast cancer, we first sequenced mtDNA from 86 cancer biopsies and available matched 50 baseline cancer-free whole blood samples from the same individuals selected from our WHILA cohort, to identify candidate mutations. Then, we designed a nested case-control study to validate if the identified mutations could be traced back to plasma samples prior to breast cancer diagnosis.

When comparing all the mtDNA mutations identified from biopsies and whole blood samples, we found that the majority of the homoplasmic and heteroplasmic mtDNA mutations were shared between the two groups, suggesting that those mutations were more likely to be germline mutations (Figure. 15A). No statistical difference was observed in the frequencies of homoplasmic mutations between biopsies and whole blood samples when stratified into different mitochondrial genome region. However, in certain regions such as D-loop, *RNR2*, *COX1*, and *CYTB*, biopsies exhibited a higher abundance of heteroplasmic mutations compared to whole blood samples (Figure. 15B-E).

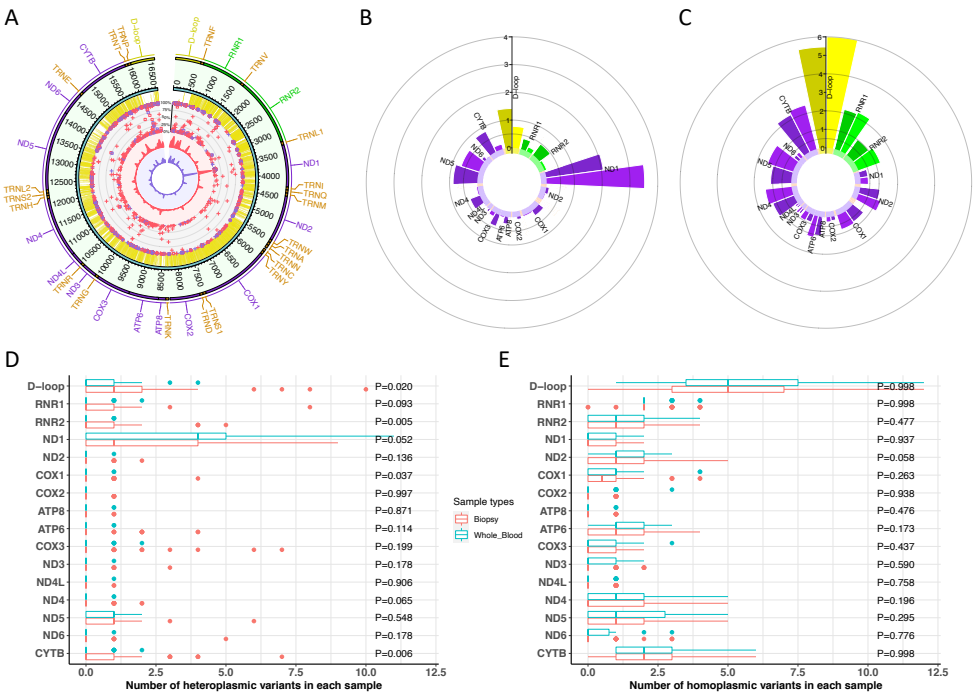


Figure 15. Mitochondrial genomes mutational landscape of 86 biopsies and available matched 50 whole blood samples.

A. Circos plot depicts the mitochondrial mutations identified in biopsies and whole blood samples. From outside of the circle to inside: (1) mitochondrial genes colored based on genome region: yellow for D-loop, purple for coding genes, green for rRNAs, orange for tRNAs. (2) mtDNA position. (3) phastCons100 conservation scores

from UCSC. (4) identified mutations with the radial axis representing the heteroplasmy level from inner 0% to outer 100%. Red crosses represent biopsies, while blues are whole blood samples. Mutations observed in both samples from the same individual are denoted with squares. (5) Mutation density with adjustment for comparison between two groups. a) red: biopsies. b) blue: whole blood samples. B-C. The mutation rate of mtDNA genomic regions. The vertical axes indicate the number of refined heteroplasmic (B) or homoplasmic (C) mutations in each gene divided by the number of samples. Darker color represents biopsies while lighter color is the whole blood samples. Mitochondrial genes are displayed and colored based on the regions as described in A. D-E. Comparison between two groups regarding mutations in mitochondrial genes.

We further analyzed the 50 matched biopsy-whole blood pairs and observed cancer biopsies had elevated occurrences of C to A/G/T and T to G heteroplasmic substitutions, while lower T to A substitutions. No statistical difference was observed in homoplasmic mutational spectra between the two groups (Figure. 16A-B). These findings suggest cancer biopsies were potentially susceptible to spontaneous deamination of cytosine. Through comparing within two groups, we categorized the mutation into three types: germline mutation (mutation present in both groups, *de novo* mutation (mutation only present in biopsies) and lost mutation (mutation only present in whole blood samples). As shown in Figure. 16C, *de novo* mutations presented in cancer biopsies had significantly higher heteroplasmy levels (HL). Subsequently, we annotated the mutations into nonsynonymous and synonymous according to the influence of the mutation on the mitochondrial genome. We found that the mutations identified in this study had a higher ratio of dN/dS. The overall dN/dS ratio was higher for heteroplasmic mutations compared to homoplasmic mutations, while *de novo* and lost heteroplasmic mutations had lower dN/dS ratios than germline mutations (Figure. 16D). Moreover, compared to matched whole blood samples, the mutations in the D-loop and rRNA genes had higher HL in biopsies (Figure. 16E). These intriguing results suggest that breast cancers manifest distinct heteroplasmic mutation signatures during cancer pathogenesis.

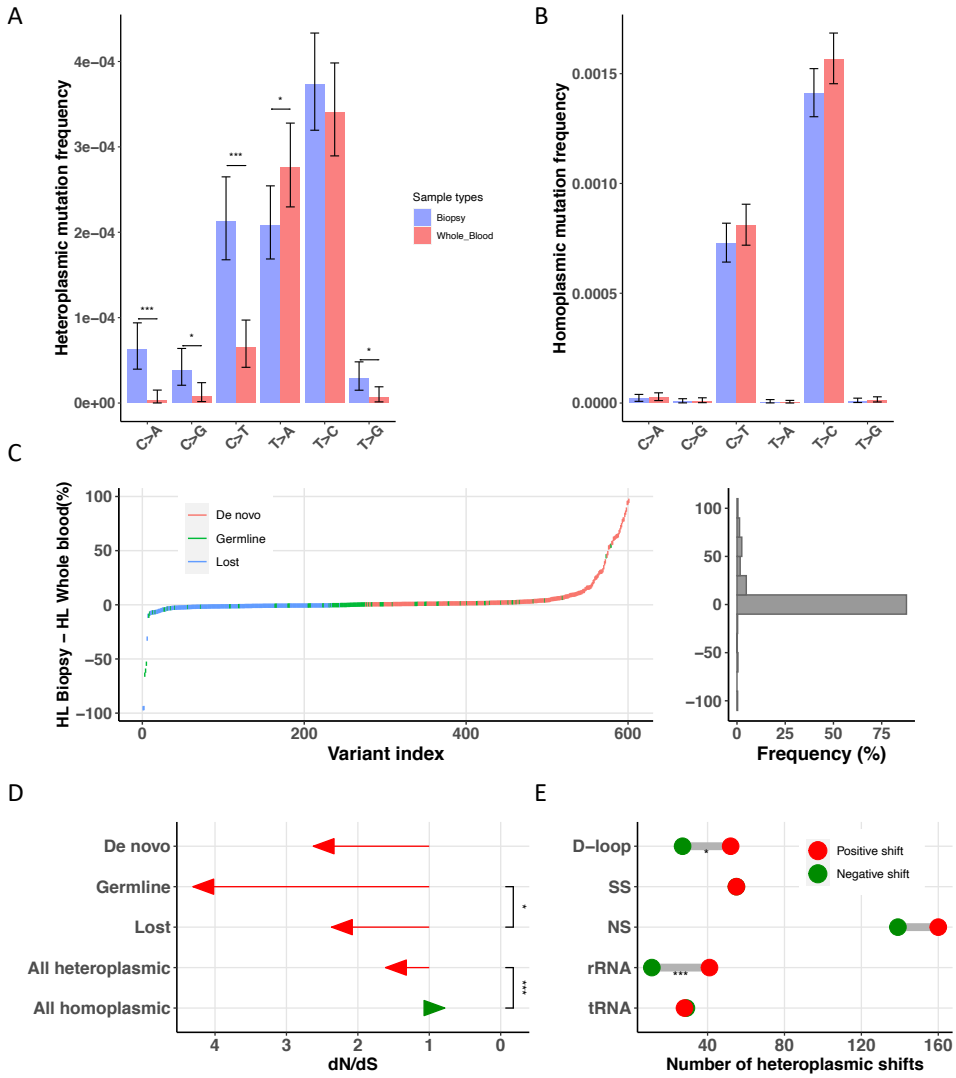


Figure 16. The positive selection of breast cancer during mitochondrial heteroplasmic shift.
A-B. The mtDNA heteroplasmic (A) and homoplasmic (B) nucleotide mutational spectra in matched 50 biopsy-whole blood pairs. C. Left, the difference in the percentage shift of HL between biopsy and the matched whole blood samples (HL Biopsy − HL whole blood) is ordered by the degree of shift. Right, the distribution of the percentage shift difference of HL between biopsy and the matched whole blood sample. D. dN/dS for all types of muatations. E. The number of heteroplasmies showing an increased or decreased HL in each mtDNA region in biopsy-whole blood pairs.

After analyses of mutation identified by NGS in two groups, we shortlisted two candidate mutations for ddPCR validation. The HL of the mtDNA mutations – mt.1888G>A and mt.16093T>C were quantified on 663 plasma samples collected before the clinical diagnosis of breast cancer. As shown in Figure 17, we

dichotomized the mutations as wild-type (HL < 10%) and mutated (HL ≥ 10%) based on the natural cut-off.

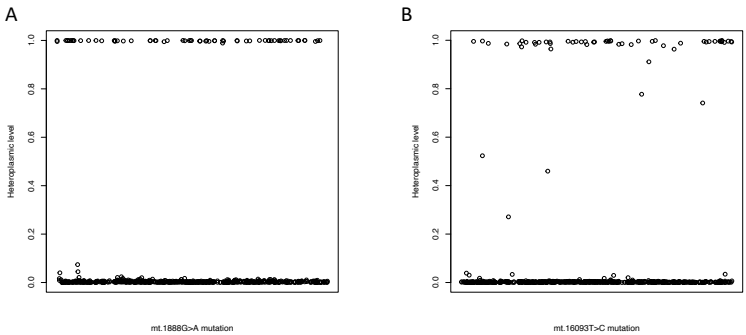


Figure 17. The HL distribution of mt.1888G>A and mt.16093T>C mutation in baseline plasma samples.

To investigate the association between the mutations and breast cancer risk, we conducted logistic regression analyses. The results showed that the presence of mt.16093T>C was significantly associated with an increased risk of breast cancer, suggesting that this mutation could potentially be used for earlier detection of breast cancer (Table 4).

Table 4. Odd ratios and 95% confidence intervals of cancer incidence associated with mtDNA mutations.

Mutation	No. of cases	No. of controls	Crude HR (95% CI)	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]
mt.1888G>A							
Heteroplasmy level	-	-	0.79 (0.37-1.21)	0.79 (0.37-1.21)	0.76 (0.34-1.19)	0.74 (0.32-1.17)	0.76 (0.33-1.19)
Dichotomized							
Wild type	280	320	1 (Ref)	1 (Ref)	1 (Ref)	1 (Ref)	1 (Ref)
Mutated	24	39	0.80 (0.38-1.21)	0.79 (0.37-1.21)	0.77 (0.35-1.19)	0.75 (0.32-1.17)	0.76 (0.33-1.19)
mt.16093T>C							
Heteroplasmy level	-	-	1.50 (1.11-1.88)	1.50 (1.12-1.88)	1.49 (1.10-1.87)	1.55 (1.16-1.93)	1.53 (1.14-1.92)
Dichotomized							
Wild type	272	335	1 (Ref)	1 (Ref)	1 (Ref)	1 (Ref)	1 (Ref)
Mutated	32	24	1.48 (1.12-1.85)	1.48 (1.12-1.85)	1.47 (1.10-1.84)	1.53 (1.15-1.90)	1.50 (1.13-1.88)

[a] Adjusted for age, BMI
[b] Adjusted for age, BMI, education level, smoking habits, alcohol consumption
[c] Adjusted for age, BMI, education level, smoking habits, alcohol consumption, activity at work, activity at home
[d] Adjusted for age, BMI, education level, smoking habits, alcohol consumption, activity at work, activity at home, diabetes, hypertension, family history of cancer

The clinical methods used today will not detect breast cancer tumors before they are visible on currently used imaging methods, however, cancer-associated molecular changes appear earlier prior to clinical diagnosis [153]. Cancer-related information is shed into the bloodstream during tumorigenesis, which potentially can be detected earlier for improving breast cancer diagnostics and patient outcomes. However, detection of early breast cancer at the molecular stage is challenging due to low

circulating copies of nuclear DNA in the blood [154]. MtDNA exists two to four orders of magnitude more copies per cell, resulting in higher abundance in the circulation and hence providing a better opportunity for earlier cancer detection [155,156]. During the preparation of this study, another mtDNA sequencing study was published comparing breast cancer biopsies and matched peripheral blood (collected before mastectomy) in a Mexican population, which showed that the mutation identified in the biopsies could be detected in the blood [157]. However, unlike the published study, our study aimed to detect and identify cancer-associated mtDNA mutations earlier than the clinical diagnosis. Thus, we included the breast biopsies, matched whole blood and plasma samples (baseline cancer-free) from the same individuals. The mtDNA mutations presented in the whole blood samples worked as a reference to rule out germline or haplogroup-associated mutations and identify the potential cancer-associated mutations. The plasma samples were used for validation of the shortlisted candidate mutations. This hypothesis-generating study demonstrated that mtDNA mutations detected in the plasma samples could potentially be used as early diagnostic biomarkers by first, comprehensively depicting the mtDNA mutation landscapes of baseline cancer-free whole blood samples and matched breast cancer biopsies, and then validating the identified candidate mutations in baseline plasma samples prior to breast cancer diagnosis.

Somatic mtDNA mutations were predominantly found to be heteroplasmic and will cause disease when exceeded a certain threshold [68]. In this study, we set up strict criteria for defining the mutations as heteroplasmic or homoplasmic and observed the positive evolutionary selection of both heteroplasmic and homoplasmic mutations with breast cancer development. Through comparing the mtDNA mutations of breast cancer biopsy with matched baseline whole blood samples from the same individual, the mutations were well-categorized according to the distribution into germline, lost and *de novo* mutations. The majority of *de novo* mutations with were under positive selection, especially in D-loop, suggested that mitochondrial transcription was potentially altered during breast cancer pathogenesis. Furthermore, a higher frequency of C to T transitions in mtDNA is a common evolutionary phenomenon. However, we observed substitution from C to other nucleotides besides T in our breast biopsies, which also suggests a potential association of the substitution with breast cancer. In this study, due to the limitations of obtaining matched whole blood samples for all biopsies included in this study, we only sequenced 50 whole blood samples and compared them with 86 cancer biopsies, preventing the identification of additional *de novo* mutations. Still, we could validate the identified mutations by ddPCR with unprecedented resolution than next-generation sequencing in the plasma samples of our breast cancer cohort. Even with a small sample size and limited statistical power to draw robust conclusions, we interpreted the findings of this study objectively and identified one heteroplasmic mutation – mt.16093T>C, the presence of which increases the risk of developing breast cancer and could be used as an early diagnostic biomarker.

Paper IV

To investigate if the mitochondrial-related gene expression has prognostic value in early-stage breast cancer, the SCANB mRNA expression dataset (n = 6657) was employed and randomly divided into a training cohort (80%, n = 5326) and a testing cohort (20%, n = 1331), TCGA-BRCA (n = 1093) worked as a validation cohort. Lasso Cox regression was conducted to build a mitochondrial signature model to predict breast cancer survival.

For this study, a 14 genes mitochondrial signature model was built and the risk score of each patient was calculated by the formula:

Mitochondrial signature risk score = –0.01855457 × *ABAT* exp. – 0.10642093 × *ABCA9* exp. + 0.00705353 × *CLPB* exp. + 0.04435106 × *MRPS9* exp. + 0.01909035 × *MTCH1* exp. + 0.05581194 × *MTHFD1L* exp. – 0.01490574 × *NAGS* exp. + 0.00607169 × *NDUFB4* exp. – 0.11098209 × *PABPC5* exp. – 0.11065619 × *PLGRKT* exp. – 0.00465041 × *RBFA* exp. + 0.038724433 × *RTN4IP1* exp. – 0.00433475 × *SLC25A24* exp. + 0.00842707 × *UQCRFS1* exp.

According to the median value of the mitochondrial signature risk score, we could divide the patients in study cohorts into low and high-risk groups. Our mitochondrial signature model demonstrated that patients with high mitochondrial signature expression had shorter overall survival (Figure. 18A-D) in the SCANB training (HR: 2.41, 95% CI: 2.08, 2.80), SCANB testing (HR: 1.77, 95% CI: 1.33, 2.36), whole SCANB cohort (HR: 2.24, 95% CI: 1.96, 2.55) as well as TCGA-BRCA validation cohort (HR:1.49, 95% CI: 1.08, 2.06). The prognostic values of mitochondrial signature were measured by the area under the receiver operating characteristic (ROC) curve (AUC) at 1-, 3- and 5-year of overall survival in our study cohorts. It yielded AUC values of 0.74, 0.70, and 0.68 for the SCANB training cohort, 0.67, 0.68, and 0.61 for the SCANB testing cohort, 0.72, 0.69, and 0.66 for the whole SCANB cohort, and 0.66, 0.63, and 0.57 for the TCGA-BRCA validation cohort, respectively (Figure. 18E-H).

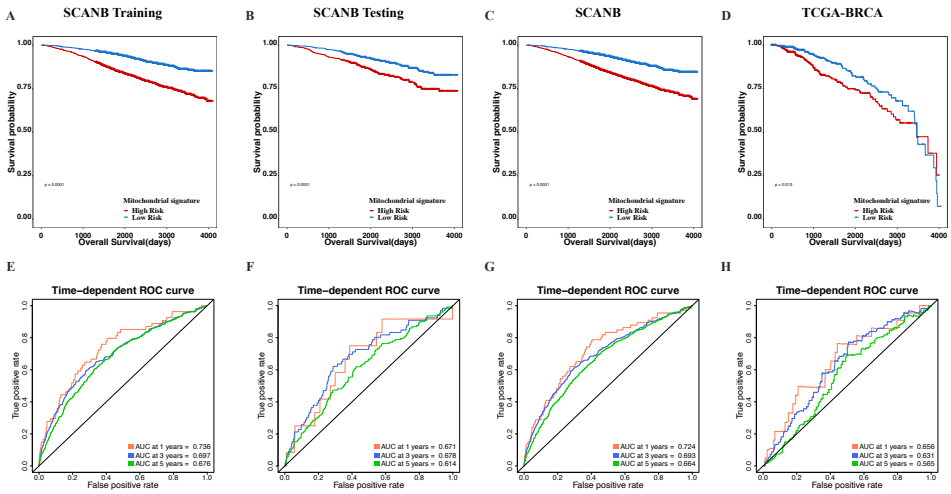


Figure 18. Evaluation of the predictive performance of the mitochondrial signature model. (A-D). Kaplan-Meier plots depict overall survival based on the expression of mitochondrial signature in breast cancer patients across the cohorts. (E-H). Time-dependent AUC of mitochondrial signature at 1-, 3-, and 5-year intervals across the cohorts.

To improve the predictive performance, we further constructed a nomogram by incorporating the mitochondrial signature with clinical variables. Multivariable Cox and stepwise regression analyses were conducted to select the most effective combination of predictors to build the nomogram. The resultant nomogram model included age, N stage, T stage, PAM50 subtypes, treatment, and the mitochondrial signature with good agreements in predicting the 1-, 3-, and 5-year survival rates. (Figure. 19A-B). Compared to other predictors, the nomogram model performed the best followed by age and mitochondrial signature (Figure. 19C-D). The prognostic values of nomogram were also measured by ROC curve at 1-, 3- and 5-year of overall survival in SCANB and TCGA-BRCA cohort. It yielded AUC values of 0.84, 0.79, and 0.79 for the whole SCANB cohort, and 0.92, 0.83, and 0.78 for the TCGA-BRCA validation cohort, respectively (Figure. 19E-F). The results demonstrated that the nomogram model exhibited high predictive accuracy in our study cohort.

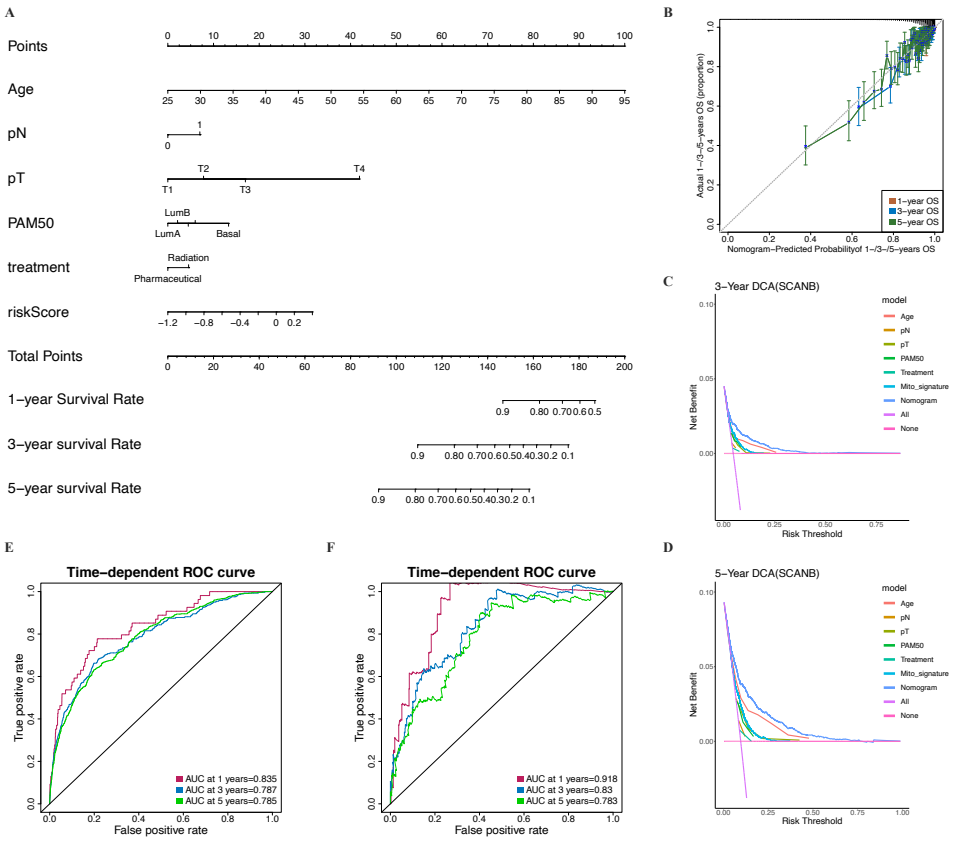


Figure 19. Establishment and evaluation of a nomogram model for breast cancer overall survival prediction.

A). Nomogram model constructed for overall survival prediction in breast cancer. pN: N stage, pT: T stage, riskScore: score calculated using mitochondrial signature formula (B). Calibration plot demonstrates the accuracy of the nomogram in the SCANB cohort. (C-D) Decision curve analysis (DCA) of nomogram performance for 3- (C), and 5-year (D) overall survival. (E-F). Time-dependent AUC analysis of nomogram model at 1-, 3-, and 5-year intervals in SCANB cohorts (E) and TCGA-BRCA cohort (F).

Early detection and accurate breast cancer diagnosis remain crucial for tailoring treatment decisions, identifying high-risk subgroups, monitoring long-term outcomes, and advancing research and patient care. Prognostic models have been developed but failed to be extensively applied widely due to the limited sample sizes and insufficient validation in external cohorts [158-165]. As we mentioned in other studies, mitochondrial dysfunction plays a crucial role in cancer development, often linked to increased aggressiveness and metastasis. Most of the genes that regulate mitochondrial function are from the nuclear genome and imported into mitochondria, the alternations in these genes may contribute to cancer pathophysiology. However, there is limited knowledge currently about the role of these genes in the overall survival of early-stage breast cancer patients. This is the

first study specifically including all known mitochondrial-related genes to build a potential prognostic model for early-stage breast cancer patients. The 14 genes mitochondrial signature model was demonstrated to be an accurate predictor of breast cancer survival and when integrated with clinical variables offered a robust prediction of clinical prognosis.

In this study, we built the model based on SCANB dataset, which is the largest early-stage breast cancer mRNA sequencing cohort originating from Sweden, and TCGA-BRCA as an external validation cohort. The large-scale dataset allowed for an unbiased analysis when employing Lasso Cox regression to automatically select relevant features and prevent the model from overfitting. However, compared to TCGA-BRCA, patients from SCANB benefitted from the free mammogram screening program and adjuvant target therapy, thus were diagnosed relatively earlier with better treatment, which influenced the survival rates and possibly reduced the prediction accuracy in the validation cohort. Thus, large and updated cohorts are needed as external validation for future perspectives. When it comes to building mRNA expression signature models using machine learning, many studies performed the feature selection based on the differential expression of the genes between two groups. In our study, we included all mitochondrial-related genes and assumed that genes differentially expressed compared to normal tissue have the potential to be prognostic biomarkers. The 14 genes we integrated into mitochondrial signature showed robust predictive value for breast cancer survival, however, the functions of each gene were not well-defined, suggesting that further studies are needed to investigate the mechanism behind these genes.

Conclusions and future perspectives

Conclusion

In conclusion, this thesis has effectively investigated the feasibility and efficacy of applying mitochondria and its related nuclear genes as potential diagnosis and prognosis biomarkers for cancer. Below is the specific conclusion for each study.

Paper I

The mitochondrial DNA copy number (mtDNA-CN), a proxy for mitochondrial dysfunction, quantified in the peripheral blood was associated with a higher risk of prevalent cancer in a cancer-specific manner. Moreover, the mtDNA-CN has the potential clinical applications as a prognostic biomarker for mortality of genital cancer.

Paper II

This data-driven Mendelian randomization study provided evidence for genetic determinants of mitochondrial dysfunction associated with the risk of cancer in a cancer-specific manner, underscoring the pivotal role of mitochondrial dysfunction in the pathogenesis of multiple cancer types. Most importantly, the identified putative causal genes can be used as potential predictive biomarkers as well as pharmacological targets for cancer prevention.

Paper III

We conducted a comprehensive characterization of the mtDNA mutation landscape in breast cancer biopsies and matched baseline whole blood samples, providing insights into the evolutionary positive selection of cancer-related heteroplasmic mtDNA mutations. Notably, the mitochondrial mutations identified in biopsies can be detected in matched plasma samples prior to cancer diagnosis, suggested that the

potential application of these mtDNA mutations as early diagnostic biomarkers for breast cancer.

Paper IV

We built a novel 14 genes mitochondrial signature model that could be an independent prognostic predictor for the survival outcomes of early-stage breast cancer. Moreover, we proposed that the nomogram model by integrating mitochondrial signature with clinical variables improved the predictive accuracy of overall early-stage breast cancer survival.

Future perspectives

The findings from this thesis have shed light on the important scientific and clinical relevance of mitochondria in cancer diagnosis and prognosis. The exploration of mitochondria and their related nuclear genes holds great promise for future cancer research and personalized medicine. Moving forward, there are several exciting perspectives to consider:

1. Potential clinical applications of mtDNA-CN: The conclusions drawn from our and other studies show heterogeneity in the association of mtDNA-CN with cancer incidence and mortality. Further research could therefore focus on improving study design sophisticatedly, standardizing mtDNA-CN measurements and validating those conclusions in larger and diverse patient cohorts to enable the clinical application of mtDNA-CN for cancer diagnosis and risk stratification.

2. Targeting mitochondrial dysfunction for cancer prevention: Mendelian randomization proved the causal effects of mitochondrial dysfunction on cancers. Future studies could investigate the mechanistic links among these causal genes, mitochondrial dysfunction and carcinogenesis, with a particular focus on the pharmacological manipulation of the identified causal genes for cancer prevention.

3. Early diagnosis of breast cancer using mtDNA mutations: We showed that mtDNA mutations could potentially be early diagnostic biomarkers for breast cancer. Prospective studies could further validate these findings and include a larger sample size to identify and detect additional cancer-associated mtDNA mutations as non-invasive and cost-effective screening tools for early breast cancer detection.

4. Mitochondrial signature for breast cancer prognosis: Our mitochondrial signature model holds great potential as an independent prognostic predictor for early-stage breast cancer outcome. Future validation using large external validation could

further refine the model, potentially leading to the implementation of this model into clinical practice to aid in patient risk stratification and personalized treatment decisions. The mechanistic function of the genes included in the model needs further investigation.

Overall, the studies conducted in this thesis provide a basis for the future application of mitochondria and its related genes as tools for cancer diagnosis, prognosis, and treatment. The identified genes and mtDNA mutations offer exciting opportunities for future scientific research, clinical application and development of drugs for cancer prevention and treatment. Continued investigation of mitochondrial biology and genetics in cancer will undoubtedly contribute to the advancement in the fight against cancer.

Acknowledgements

In embarking upon the culmination of my PhD, I have been privileged to receive guidance, support, and encouragement from numerous individuals that have played instrumental roles in shaping the trajectory of my research journey. The completion of this doctoral thesis stands as a testament not only to my dedication but also to the unwavering contributions of those who have helped me throughout the last four years.

First, I would like to extend my heartfelt gratitude to my supervisor, **Ashfaque Memon**, for your guidance throughout the journey of completing this doctoral thesis. Your expertise, patience, insightful feedback, constructive criticism, and thought-provoking discussions have been pivotal in improving the content, direction and scientific presentation both orally and in writing of my PhD research. I am profoundly grateful to you for your help, support and for the academic freedom that you offered.

My gratitude extends to my co-supervisor **Kristina Sundquist**, thank you for having me as a PhD student in your brilliant team. You are such a great mentor and leader with a strong heart, attentive ears, incredible wisdom and kindness. I am truly honored to have had the privilege of working in CPF with your unwavering support and guidance even during your demanding schedule. I will always remember your trust and encouragement, and draw strength from the lesson that life has taught me.

I would also like to express my grateful appreciation to my co-supervisor, **Xiao Wang**, for your invaluable help and support that extend beyond the academic realm. Your enthusiasm and passion for both research and life have inspired me and also broadened my perspectives. I am truly fortunate to have you as a co-supervisor, and I acknowledge you for your exceptional role in shaping both my academic and personal growth.

I extend my deepest thanks to **Anna Hedelius**, you are a sweet, responsible and wonderful person. Your expertise in teaching intricate techniques and patient instruction in the usage of all machines in the lab was crucial to the success of my research. Your consistent warmth and caring have provided me with unwavering support. I consider myself genuinely fortunate to have been able to learn and grow

under your gmentorship. Furthermore, your skill for baking the most exquisite cakes and in selecting the perfect gifts for us have never ceased to amaze.

Thank you, **Sakshi Vats**, my sister from another mother and my cherished friend. Throughout the challenging phases of research and writing, you provided valuable insights, encouragement, and a listening ear that made a significant difference. Your willingness to share knowledge and exchange ideas has a positive impact on both my academic and personal life. Thank you for your friendship and kindness, I extend my heartfelt appreciation for your enduring presence by my side and am starting to miss you before finishing this thesis.

My heartfelt thanks go to **Jan Sundquist**, the former director of CPF, for providing us with an amazing platform to do research. Thank you for your invaluable support and intriguing discussion at our Wednesday group meeting.

I want to express my great thanks to the family members of our CPF lab, **Lubna Memon**, **Jianguang Ji**, **Mattias Hedelius** and **Bhavik Rathod**, for your generous hospitality and unwavering kindness that made my doctoral journey exceptionally rewarding.

My special thanks go to **Patrick Reilly**, for your invaluable assistance in polishing my writing, refining and enhancing the quality of my manuscripts and this thesis. Your dedication will always be remembered with deep gratitude.

Karolina Palmér and **Kenta Okuyama**, thank you for organizing the epi-seminars and metho-seminars. These seminars not only deepened my understanding of statistics but also broadened my perspective.

Thank you, **Mats-Åke Persson** and **Helene Brandt**, your swift assistance in resolving technical issues and ordering electronic components is greatly appreciated.

Wuqing Huang, **Yishan Liu**, **Sanjay Thompson** and **Wazah Pello Esso**, thank you my fellow PhD colleagues for engaging in captivating discussions about our respective projects and for generously sharing insights about our personal lives.

I extend my gratitude to my colleagues at CPF for engaging in insightful discussions during our staff day and fika. Your presence and active participation have played a significant role in enhancing the overall enjoyment of working at CPF.

I would like to thank **Sinh Tran**, for your incredible generosity in gifting me plants each year, which has greatly contributed to the flourishing of my balcony. Additionally, I want to express my appreciation for your willingness to share the antibodies and disposable materials.

I would like to express my sincere gratitude to my colleagues on the 2nd and 4th floors of the Wallenberg building. My heartfelt thanks extend to all the researchers who actively participate in the engaging Wednesday cake club and the insightful Thursday morning seminars. Your willingness to share and discuss your remarkable research has greatly enriched our academic environment.

Hong Luo, **Yueli Yao** and **Jiesheng Lin**, thank you for being such wonderful companions during our travels. Your friendship has truly enriched our journeys, and I am incredibly thankful for the remarkable moments we shared while exploring the world and immersing ourselves in the breathtaking landscapes.

Jingxue Pan, **Mi Huang**, **Ming Sun, Guoqiao Zheng** and **Subhayan Chattopadhyay**, thank you for your incredible friendship and the cherished moments we've shared. The time we've spent together, whether it was fishing, enjoying dinners and engaging in lively table games or coming together to celebrate Chinese traditional holidays, were treasured memories that I will always hold dear.

Rui Wu, **Zhiyi Ding**, **Shu Yang**, **Cheng Luan**, **Jiangming Sun**, **Yan Borné, Mengyu Pan** and **Baojun Zhong**, thank you for being friends and lunch companions. Your delightful presence and engaging conversations significantly illuminate and enrich our workdays, creating a truly uplifting and enjoyable atmosphere.

My little furry guy, **Hugo**, your unwavering presence provided immeasurable joy throughout the journey of completing this doctoral thesis. Your boundless enthusiasm and unconditional affection served as a constant source of inspiration, reminding me to embrace moments of happiness through challenging times. The comfort and companionship you offered were a reminder of the simple joys in life, and for that, I am truly thankful. My furry friend's contribution, though unspoken, was undeniably significant, and I am honored to acknowledge his role in this endeavor.

Naiqi Zhang, over the past four years, the majority of my time has been shared alongside you. Throughout this time, I've held the firm belief that I could not encounter a more righteous and loyal friend or dedicated family member than you. You have been taking care of me and supporting me mentally and statistically :). Your boundless patience, encouragement, and unending belief in my abilities have shaped my personality and transformed me into a better person equipped for effective communication. Through your presence, the universe seems illuminated when the weight of academia became overwhelming, and joy intensified when you celebrated with me in moments of accomplishment. Thank you for being my best friend, family and my great source of strength.

In the end, I want to express my sincerest and heartfelt appreciation to my parents, for the unwavering support, continuous encouragement and immeasurable love you have provided throughout the entirety of my research journey. I am truly grateful for the way you have consistently respected my decisions and choices, allowing me the space to grow and explore. Your love, sacrifices and constant belief in my abilities have been a cornerstone of my motivation and my pillars of strength, guiding me through the challenges and triumphs of this academic pursuit. This accomplishment would not have been possible without your boundless care and guidance. Thank you, Mom and Dad, for the love and encouragement, which have molded me into the person I am today, and I am profoundly grateful for everything you have done for me.

References

1. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov 2022;12(1):31-46.
2. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71(3):209-249.
3. Chen S, Cao Z, Prettner K, et al. Estimates and Projections of the Global Economic Cost of 29 Cancers in 204 Countries and Territories From 2020 to 2050. JAMA Oncol 2023;9(4):465-472.
4. Turnbull C, Sud A, Houlston RS. Cancer genetics, precision prevention and a call to action. Nat Genet 2018;50(9):1212-1218.
5. Garber JE, Offit K. Hereditary cancer predisposition syndromes. J Clin Oncol 2005;23(2):276-92.
6. Frank SA. Genetic predisposition to cancer - insights from population genetics. Nat Rev Genet 2004;5(10):764-72.
7. Fletcher O, Houlston RS. Architecture of inherited susceptibility to common cancer. Nat Rev Cancer 2010;10(5):353-61.
8. Chakravarty D, Solit DB. Clinical cancer genomic profiling. Nat Rev Genet 2021;22(8):483-501.
9. Weitzel JN, Blazer KR, MacDonald DJ, Culver JO, Offit K. Genetics, genomics, and cancer risk assessment: State of the Art and Future Directions in the Era of Personalized Medicine. CA Cancer J Clin 2011;61(5):327-59.
10. Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. Nat Rev Cancer 2017;17(11):692-704.
11. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nat Genet 2003;33 Suppl:238-44.
12. Yang X, Kar S, Antoniou AC, Pharoah PDP. Polygenic scores in cancer. Nat Rev Cancer 2023.
13. Vineis P, Wild CP. Global cancer patterns: causes and prevention. Lancet 2014;383(9916):549-57.
14. Malvezzi M, Bosetti C, Rosso T, et al. Lung cancer mortality in European men: trends and predictions. Lung Cancer 2013;80(2):138-45.
15. Hendriks HFJ. Alcohol and Human Health: What Is the Evidence? Annu Rev Food Sci Technol 2020;11:1-21.
16. Rumgay H, Murphy N, Ferrari P, Soerjomataram I. Alcohol and Cancer: Epidemiology and Biological Mechanisms. Nutrients 2021;13(9).
17. Mayne ST, Playdon MC, Rock CL. Diet, nutrition, and cancer: past, present and future. Nat Rev Clin Oncol 2016;13(8):504-15.

18. Huang Y, Cao D, Chen Z, et al. Red and processed meat consumption and cancer outcomes: Umbrella review. Food Chem 2021;356:129697.
19. Kerr J, Anderson C, Lippman SM. Physical activity, sedentary behaviour, diet, and cancer: an update and emerging new evidence. Lancet Oncol 2017;18(8):e457-e471.
20. Gu KJ, Li G. An Overview of Cancer Prevention: Chemoprevention and Immunoprevention. J Cancer Prev 2020;25(3):127-135.
21. Collatuzzo G, Boffetta P. Cancers Attributable to Modifiable Risk Factors: A Road Map for Prevention. Annu Rev Public Health 2023;44:279-300.
22. Rigel DS. Cutaneous ultraviolet exposure and its relationship to the development of skin cancer. J Am Acad Dermatol 2008;58(5 Suppl 2):S129-32.
23. Lumniczky K, Impens N, Armengol G, et al. Low dose ionizing radiation effects on the immune system. Environ Int 2021;149:106212.
24. Ke Y, Wang XJ. TGFbeta Signaling in Photoaging and UV-Induced Skin Cancer. J Invest Dermatol 2021;141(4S):1104-1110.
25. Crosby D, Bhatia S, Brindle KM, et al. Early detection of cancer. Science 2022;375(6586):eaay9040.
26. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer 2005;5(11):845-56.
27. Badowski C, He B, Garmire LX. Blood-derived lncRNAs as biomarkers for cancer diagnosis: the Good, the Bad and the Beauty. NPJ Precis Oncol 2022;6(1):40.
28. Arrieta O, Cacho B, Morales-Espinosa D, Ruelas-Villavicencio A, Flores-Estrada D, Hernandez-Pedro N. The progressive elevation of alpha fetoprotein for the diagnosis of hepatocellular carcinoma in patients with liver cirrhosis. BMC Cancer 2007;7:28.
29. Zamay TN, Zamay GS, Kolovskaya OS, et al. Current and Prospective Protein Biomarkers of Lung Cancer. Cancers (Basel) 2017;9(11).
30. Sardana G, Dowell B, Diamandis EP. Emerging biomarkers for the diagnosis and prognosis of prostate cancer. Clin Chem 2008;54(12):1951-60.
31. Babayan A, Pantel K. Advances in liquid biopsy approaches for early detection and monitoring of cancer. Genome Med 2018;10(1):21.
32. Lone SN, Nisar S, Masoodi T, et al. Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments. Mol Cancer 2022;21(1):79.
33. Xie W, Suryaprakash S, Wu C, Rodriguez A, Fraterman S. Trends in the use of liquid biopsy in oncology. Nat Rev Drug Discov 2023;22(8):612-613.
34. Yu W, Hurley J, Roberts D, et al. Exosome-based liquid biopsies in cancer: opportunities and challenges. Ann Oncol 2021;32(4):466-477.
35. Shen SY, Singhania R, Fehringer G, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature 2018;563(7732):579-583.
36. Cristiano S, Leal A, Phallen J, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. Nature 2019;570(7761):385-389.
37. Evans A, Neuman N. The Mighty Mitochondria. Mol Cell 2016;61(5):641.
38. Atlante A, Valenti D. Mitochondria Have Made a Long Evolutionary Path from Ancient Bacteria Immigrants within Eukaryotic Cells to Essential Cellular Hosts

and Key Players in Human Health and Disease. Curr Issues Mol Biol 2023;45(5):4451-4479.
39. Warburg O, Wind F, Negelein E. The Metabolism of Tumors in the Body. J Gen Physiol 1927;8(6):519-30.
40. Vyas S, Zaganjor E, Haigis MC. Mitochondria and Cancer. Cell 2016;166(3):555-566.
41. Vasan K, Werner M, Chandel NS. Mitochondrial Metabolism as a Target for Cancer Therapy. Cell Metab 2020;32(3):341-352.
42. Sabharwal SS, Schumacker PT. Mitochondrial ROS in cancer: initiators, amplifiers or an Achilles' heel? Nat Rev Cancer 2014;14(11):709-21.
43. Wallace DC. Mitochondria and cancer. Nat Rev Cancer 2012;12(10):685-98.
44. Delaunay S, Pascual G, Feng B, et al. Mitochondrial RNA modifications shape metabolic plasticity in metastasis. Nature 2022;607(7919):593-603.
45. Bock FJ, Tait SWG. Mitochondria as multifaceted regulators of cell death. Nat Rev Mol Cell Biol 2020;21(2):85-100.
46. Harrington JS, Ryter SW, Plataki M, Price DR, Choi AMK. Mitochondria in health, disease, and aging. Physiol Rev 2023;103(4):2349-2422.
47. Marchi S, Guilbaud E, Tait SWG, Yamazaki T, Galluzzi L. Mitochondrial control of inflammation. Nat Rev Immunol 2023;23(3):159-173.
48. Amorim JA, Coppotelli G, Rolo AP, Palmeira CM, Ross JM, Sinclair DA. Mitochondrial and metabolic dysfunction in ageing and age-related diseases. Nat Rev Endocrinol 2022;18(4):243-258.
49. Kuo CL, Ponneri Babuharisankar A, Lin YC, et al. Mitochondrial oxidative stress in the tumor microenvironment and cancer immunoescape: foe or friend? J Biomed Sci 2022;29(1):74.
50. Han Y, Cho U, Kim S, et al. Tumour microenvironment on mitochondrial dynamics and chemoresistance in cancer. Free Radic Res 2018;52(11-12):1271-1287.
51. Scheid AD, Beadnell TC, Welch DR. Roles of mitochondria in the hallmarks of metastasis. Br J Cancer 2021;124(1):124-135.
52. Porporato PE, Filigheddu N, Pedro JMB, Kroemer G, Galluzzi L. Mitochondrial metabolism and cancer. Cell Res 2018;28(3):265-280.
53. Chan DC. Mitochondrial Dynamics and Its Involvement in Disease. Annu Rev Pathol 2020;15:235-259.
54. Bonnay F, Veloso A, Steinmann V, et al. Oxidative Metabolism Drives Immortalization of Neural Stem Cells during Tumorigenesis. Cell 2020;182(6):1490-1507 e19.
55. Bensard CL, Wisidagama DR, Olson KA, et al. Regulation of Tumor Initiation by the Mitochondrial Pyruvate Carrier. Cell Metab 2020;31(2):284-300 e7.
56. Ashton TM, McKenna WG, Kunz-Schughart LA, Higgins GS. Oxidative Phosphorylation as an Emerging Target in Cancer Therapy. Clin Cancer Res 2018;24(11):2482-2490.
57. Sun N, Youle RJ, Finkel T. The Mitochondrial Basis of Aging. Mol Cell 2016;61(5):654-666.

58. Moschoi R, Imbert V, Nebout M, et al. Protective mitochondrial transfer from bone marrow stromal cells to acute myeloid leukemic cells during chemotherapy. Blood 2016;128(2):253-64.
59. Burt R, Dey A, Aref S, et al. Activated stromal cells transfer mitochondria to rescue acute lymphoblastic leukemia cells from oxidative stress. Blood 2019;134(17):1415-1429.
60. Saha T, Dash C, Jayabalan R, et al. Intercellular nanotubes mediate mitochondrial trafficking between cancer and immune cells. Nat Nanotechnol 2021.
61. Baldwin JG, Gattinoni L. Cancer cells hijack T-cell mitochondria. Nat Nanotechnol 2021.
62. Floros VI, Pyle A, Dietmann S, et al. Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos. Nat Cell Biol 2018;20(2):144-151.
63. Wei W, Tuna S, Keogh MJ, et al. Germline selection shapes human mitochondrial DNA diversity. Science 2019;364(6442).
64. Li M, Schroder R, Ni S, Madea B, Stoneking M. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. Proc Natl Acad Sci U S A 2015;112(8):2491-6.
65. Hahn A, Zuryn S. The Cellular Mitochondrial Genome Landscape in Disease. Trends Cell Biol 2019;29(3):227-240.
66. Gammage PA, Frezza C. Mitochondrial DNA: the overlooked oncogenome? BMC Biol 2019;17(1):53.
67. Perez-Amado CJ, Bazan-Cordoba A, Hidalgo-Miranda A, Jimenez-Morales S. Mitochondrial Heteroplasmy Shifting as a Potential Biomarker of Cancer Progression. Int J Mol Sci 2021;22(14).
68. Yuan Y, Ju YS, Kim Y, et al. Comprehensive molecular characterization of mitochondrial genomes in human cancers. Nat Genet 2020;52(3):342-352.
69. Mi J, Tian G, Liu S, et al. The relationship between altered mitochondrial DNA copy number and cancer risk: a meta-analysis. Sci Rep 2015;5:10039.
70. Hu L, Yao X, Shen Y. Altered mitochondrial DNA copy number contributes to human cancer risk: evidence from an updated meta-analysis. Sci Rep 2016;6:35859.
71. Giaccherini M, Gentiluomo M, Fornili M, Lucenteforte E, Baglietto L, Campa D. Association between telomere length and mitochondrial copy number and cancer risk in humans: A meta-analysis on more than 300,000 individuals. Crit Rev Oncol Hematol 2021;167:103510.
72. Li Y, Sundquist K, Wang X, et al. Association of Mitochondrial DNA Copy Number and Telomere Length with Prevalent and Incident Cancer and Cancer Mortality in Women: A Prospective Swedish Population-Based Study. Cancers (Basel) 2021;13(15).
73. Chen N, Wen S, Sun X, et al. Elevated Mitochondrial DNA Copy Number in Peripheral Blood and Tissue Predict the Opposite Outcome of Cancer: A Meta-Analysis. Sci Rep 2016;6:37404.
74. Kim M, Mahmood M, Reznik E, Gammage PA. Mitochondrial DNA is a major source of driver mutations in cancer. Trends Cancer 2022;8(12):1046-1059.

75. Bai RK, Leal SM, Covarrubias D, Liu A, Wong LJ. Mitochondrial genetic background modifies breast cancer risk. Cancer Res 2007;67(10):4687-94.
76. Salas A, Garcia-Magarinos M, Logan I, Bandelt HJ. The saga of the many studies wrongly associating mitochondrial DNA with breast cancer. BMC Cancer 2014;14:659.
77. Li Y, Sundquist K, Zhang N, Wang X, Sundquist J, Memon AA. Mitochondrial related genome-wide Mendelian randomization identifies putatively causal genes for multiple cancer types. EBioMedicine 2023;88:104432.
78. Navaglia F, Basso D, Fogar P, et al. Mitochondrial DNA D-loop in pancreatic cancer: somatic mutations are epiphenomena while the germline 16519 T variant worsens metabolism and outcome. Am J Clin Pathol 2006;126(4):593-601.
79. van Gisbergen MW, Voets AM, Starmans MH, et al. How do changes in the mtDNA and mitochondrial dysfunction influence cancer and cancer therapy? Challenges, opportunities and models. Mutat Res Rev Mutat Res 2015;764:16-30.
80. Moro L. Mitochondrial DNA and MitomiR Variations in Pancreatic Cancer: Potential Diagnostic and Prognostic Biomarkers. Int J Mol Sci 2021;22(18).
81. Larman TC, DePalma SR, Hadjipanayis AG, et al. Spectrum of somatic mitochondrial mutations in five cancers. Proc Natl Acad Sci U S A 2012;109(35):14087-91.
82. Dennerlein S, Wang C, Rehling P. Plasticity of Mitochondrial Translation. Trends Cell Biol 2017;27(10):712-721.
83. Gao H, Xing F. A novel signature model based on mitochondrial-related genes for predicting survival of colon adenocarcinoma. BMC Med Inform Decis Mak 2022;22(1):277.
84. Kang J, Li N, Wang F, et al. Exploration of Reduced Mitochondrial Content-Associated Gene Signature and Immunocyte Infiltration in Colon Adenocarcinoma by an Integrated Bioinformatic Analysis. Front Genet 2022;13:832331.
85. Marquardt A, Solimando AG, Kerscher A, et al. Subgroup-Independent Mapping of Renal Cell Carcinoma-Machine Learning Reveals Prognostic Mitochondrial Gene Signature Beyond Histopathologic Boundaries. Front Oncol 2021;11:621278.
86. Wu Y, Zhang X, Wei X, et al. A Mitochondrial Dysfunction and Oxidative Stress Pathway-Based Prognostic Signature for Clear Cell Renal Cell Carcinoma. Oxid Med Cell Longev 2021;2021:9939331.
87. Xia Z, Liu H, Fan S, et al. A Novel Four Mitochondrial Respiration-Related Signature for Predicting Biochemical Recurrence of Prostate Cancer. J Clin Med 2023;12(2).
88. Zhang X, Wu H, Niu J, et al. A novel mitochondria-related gene signature in esophageal carcinoma: prognostic, immune, and therapeutic features. Funct Integr Genomics 2023;23(2):109.
89. Jiang X, Xia Y, Meng H, et al. Identification of a Nuclear Mitochondrial-Related Multi-Genes Signature to Predict the Prognosis of Bladder Cancer. Front Oncol 2021;11:746029.

90. Jin X, Zhang H, Sui Q, et al. Identification and validation of the mitochondrial function related hub genes by unsupervised machine learning and multi-omics analyses in lung adenocarcinoma. Heliyon 2022;8(12):e11966.
91. Liu J, Zhang F, Zhong J, Zheng Z. Signature and Molecular Mechanism of Mitochondrial Energy Metabolism Pathway-Related Genes in Lung Adenocarcinoma. Dis Markers 2022;2022:3201600.
92. Li YP, Liu GX, Wu ZL, et al. A Novel Mitochondrial-Related Gene Signature for the Tumor Immune Microenvironment Evaluation and Prognosis Prediction in Lung Adenocarcinoma. J Immunol Res 2022;2022:5366185.
93. Zhang X, Dong W, Zhang J, et al. A Novel Mitochondrial-Related Nuclear Gene Signature Predicts Overall Survival of Lung Adenocarcinoma Patients. Front Cell Dev Biol 2021;9:740487.
94. Lee YK, Jee BA, Kwon SM, et al. Identification of a mitochondrial defect gene signature reveals NUPR1 as a key regulator of liver cancer progression. Hepatology 2015;62(4):1174-89.
95. Wang Y, Song F, Zhang X, Yang C. Mitochondrial-Related Transcriptome Feature Correlates with Prognosis, Vascular Invasion, Tumor Microenvironment, and Treatment Response in Hepatocellular Carcinoma. Oxid Med Cell Longev 2022;2022:1592905.
96. Chang J, Wu H, Wu J, et al. Constructing a novel mitochondrial-related gene signature for evaluating the tumor immune microenvironment and predicting survival in stomach adenocarcinoma. J Transl Med 2023;21(1):191.
97. Taylor SC, Laperriere G, Germain H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. Sci Rep 2017;7(1):2409.
98. Hayden RT, Gu Z, Ingersoll J, et al. Comparison of droplet digital PCR to real-time PCR for quantitative detection of cytomegalovirus. J Clin Microbiol 2013;51(2):540-6.
99. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ 2018;362:k601.
100. Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. Int J Epidemiol 2016;45(3):908-15.
101. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16(4):385-95.
102. Samsioe G, Lidfeldt J, Nerbrand C, Nilsson P. The women's health in the Lund area (WHILA) study--an overview. Maturitas 2010;65(1):37-45.
103. 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.
104. McRae AF, Marioni RE, Shah S, et al. Identification of 55,000 Replicated DNA Methylation QTL. Sci Rep 2018;8(1):17605.
105. 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.

106. 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.
107. 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.
108. 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.
109. 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.
110. Zhang H, Ahearn TU, Lecarpentier J, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat Genet 2020;52(6):572-581.
111. Backman JD, Li AH, Marcketta A, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. Nature 2021;599(7886):628-634.
112. Lyon MS, Andrews SJ, Elsworth B, Gaunt TR, Hemani G, Marcora E. The variant call format provides efficient and robust storage of GWAS summary statistics. Genome Biol 2021;22(1):32.
113. Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet 2021;53(10):1415-1424.
114. O'Mara TA, Glubb DM, Amant F, et al. Identification of nine new susceptibility loci for endometrial cancer. Nat Commun 2018;9(1):3166.
115. Laskar RS, Muller DC, Li P, et al. Sex specific associations in genome wide association analysis of renal cell carcinoma. Eur J Hum Genet 2019;27(10):1589-1598.
116. Trepo E, Caruso S, Yang J, et al. Common genetic variation in alcohol-related hepatocellular carcinoma: a case-control genome-wide association study. Lancet Oncol 2022;23(1):161-171.
117. Wang Y, McKay JD, Rafnar T, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet 2014;46(7):736-41.
118. Lesseur C, Diergaarde B, Olshan AF, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. Nat Genet 2016;48(12):1544-1550.
119. Phelan CM, Kuchenbaecker KB, Tyrer JP, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nat Genet 2017;49(5):680-691.
120. Schumacher FR, Al Olama AA, Berndt SI, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet 2018;50(7):928-936.
121. 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 2023;4(1):100147.
122. Staaf J, Hakkinen J, Hegardt C, et al. RNA sequencing-based single sample predictors of molecular subtype and risk of recurrence for clinical assessment of early-stage breast cancer. NPJ Breast Cancer 2022;8(1):94.

123. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 2013;45(10):1113-20.
124. Memon AA, Zoller B, Hedelius A, et al. Quantification of mitochondrial DNA copy number in suspected cancer patients by a well optimized ddPCR method. Biomol Detect Quantif 2017;13:32-39.
125. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015;526(7571):68-74.
126. Wu Y, Zeng J, Zhang F, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nat Commun 2018;9(1):918.
127. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-60.
128. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol 2017;32(5):377-389.
129. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018;50(5):693-698.
130. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology 2017;28(1):30-42.
131. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res 2017;26(5):2333-2355.
132. 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.
133. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32(19):3047-8.
134. Krueger F, James F, Ewels P, et al. FelixKrueger/TrimGalore: v0.6.10 - add default decompression path (0.6.10). Zenodo. (https://zenodo.org/record/7598955).
135. Calabrese C, Simone D, Diroma MA, et al. MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. Bioinformatics 2014;30(21):3115-7.
136. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;27(21):2987-93.
137. Ruiz-Pesini E, Lott MT, Procaccio V, et al. An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Res 2007;35(Database issue):D823-8.
138. Weissensteiner H, Pacher D, Kloss-Brandstatter A, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res 2016;44(W1):W58-63.
139. 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.

140. Memon AA, Vats S, Sundquist J, Li Y, Sundquist K. Mitochondrial DNA Copy Number: Linking Diabetes and Cancer. Antioxid Redox Signal 2022;37(16-18):1168-1190.
141. Abd Radzak SM, Mohd Khair SZN, Ahmad F, Patar A, Idris Z, Mohamed Yusoff AA. Insights regarding mitochondrial DNA copy number alterations in human cancer (Review). Int J Mol Med 2022;50(2).
142. Zhang W, Lin S, Zeng B, et al. High leukocyte mitochondrial DNA copy number contributes to poor prognosis in breast cancer patients. BMC Cancer 2023;23(1):377.
143. Hosgood HD, 3rd, Davitt M, Cawthon R, et al. Mitochondrial DNA Fragmentation and Risk of Non-Hodgkin Lymphoma. JAMA Netw Open 2023;6(8):e2326885.
144. Bernal-Tirapo J, Bayo Jimenez MT, Yuste-Garcia P, et al. Evaluation of Mitochondrial Function in Blood Samples Shows Distinct Patterns in Subjects with Thyroid Carcinoma from Those with Hyperplasia. Int J Mol Sci 2023;24(7).
145. Hagg S, Jylhava J, Wang Y, Czene K, Grassmann F. Deciphering the genetic and epidemiological landscape of mitochondrial DNA abundance. Hum Genet 2021;140(6):849-861.
146. Lund M, Melbye M, Diaz LJ, Duno M, Wohlfahrt J, Vissing J. Mitochondrial dysfunction and risk of cancer. Br J Cancer 2015;112(6):1134-40.
147. 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.
148. 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):1117-23.
149. Tong XY, Liao X, Gao M, et al. Identification of NUDT5 Inhibitors From Approved Drugs. Front Mol Biosci 2020;7:44.
150. Page BDG, Valerie NCK, Wright RHG, et al. Targeted NUDT5 inhibitors block hormone signaling in breast cancer cells. Nat Commun 2018;9(1):250.
151. 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):527-39.
152. 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):15253-8.
153. Fowler AM, Strigel RM. Clinical advances in PET-MRI for breast cancer. Lancet Oncol 2022;23(1):e32-e43.
154. Alba-Bernal A, Lavado-Valenzuela R, Dominguez-Recio ME, et al. Challenges and achievements of liquid biopsy technologies employed in early breast cancer. EBioMedicine 2020;62:103100.
155. Balachander K, Roy A, Priyadharsini JV, Murugan S, Paramasivam A. Mitochondrial DNA in circulating exosomes: A novel biomarker and potential therapeutic target for oral cancer. Oral Oncol 2022;128:105857.

156. Lyu X, Tsui YM, Ho DW, Ng IO. Liquid Biopsy Using Cell-Free or Circulating Tumor DNA in the Management of Hepatocellular Carcinoma. Cell Mol Gastroenterol Hepatol 2022;13(6):1611-1624.
157. Perez-Amado CJ, Tovar H, Gomez-Romero L, et al. Mitochondrial DNA Mutation Analysis in Breast Cancer: Shifting From Germline Heteroplasmy Toward Homoplasmy in Tumors. Front Oncol 2020;10:572954.
158. Tao C, Luo R, Song J, Zhang W, Ran L. A seven-DNA methylation signature as a novel prognostic biomarker in breast cancer. J Cell Biochem 2020;121(3):2385-2393.
159. Zhang Y, Li Z, Chen M, et al. Identification of a New Eight-Long Noncoding RNA Molecular Signature for Breast Cancer Survival Prediction. DNA Cell Biol 2019;38(12):1529-1539.
160. Kang S, Kim EH, Hwang JE, et al. Prognostic significance of high metabolic activity in breast cancer: PET signature in breast cancer. Biochem Biophys Res Commun 2019;511(1):185-191.
161. Junjun S, Yangyanqiu W, Jing Z, et al. Prognostic model based on six PD-1 expression and immune infiltration-associated genes predicts survival in breast cancer. Breast Cancer 2022;29(4):666-676.
162. Wang D, Wei G, Ma J, et al. Identification of the prognostic value of ferroptosis-related gene signature in breast cancer patients. BMC Cancer 2021;21(1):645.
163. Zhang D, Duan Y, Cun J, Yang Q. Identification of Prognostic Alternative Splicing Signature in Breast Carcinoma. Front Genet 2019;10:278.
164. Li J, Wu F, Li C, et al. The cuproptosis-related signature predicts prognosis and indicates immune microenvironment in breast cancer. Front Genet 2022;13:977322.
165. Pei S, Zhang P, Chen H, et al. Integrating single-cell RNA-seq and bulk RNA-seq to construct prognostic signatures to explore the role of glutamine metabolism in breast cancer. Front Endocrinol (Lausanne) 2023;14:1135297.

Paper I

cancers

MDPI

Article

Association of Mitochondrial DNA Copy Number and Telomere Length with Prevalent and Incident Cancer and Cancer Mortality in Women: A Prospective Swedish Population-Based Study

Yanni Li [1,*], Kristina Sundquist [1,2,3], Xiao Wang [1], Naiqi Zhang [1], Anna Hedelius [1], Jan Sundquist [1,2,3] and Ashfaque A. Memon [1]

[1] Center for Primary Health Care Research, Lund University, 20502 Malmö, Sweden
[2] Department of Family Medicine and Community Health, Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[3] Center for Community-Based Healthcare Research and Education (CoHRE), Department of Functional Pathology, School of Medicine, Shimane University, Izumo, Shimane 693-8501, Japan
* Correspondence: yanni.li@med.lu.se

Simple Summary: Individuals with abnormal alterations in mitochondrial DNA copy number (mtDNA-CN) and telomere length are at higher risk of developing certain types of cancer. This report suggests that mtDNA-CN and relative telomere length measured in peripheral blood have potential clinical applications for risk prediction of different cancers and that mtDNA-CN could be used as a prognostic biomarker in malignancy. This comprehensive work strengthens several previous relevant findings in certain types of cancer and broadens our understanding of the link between mtDNA-CN, telomere length and future risk of many cancer types. The translational implication of our findings is that postmenopausal genital organ cancer patients with lower levels of baseline mtDNA-CN or shorter telomere length can be identified for early adjustment of lifestyle and hormone replacement therapy.

Abstract: Changes in mitochondrial DNA copy number (mtDNA-CN) and telomere length have, separately, been proposed as risk factors for various cancer types. However, those results are conflicting. Here, mtDNA-CN and relative telomere length were measured in 3225 middle-aged women included in a large population-based prospective cohort. The baseline mtDNA-CN in patients with prevalent breast cancer was significantly higher (12.39 copies/μL) than cancer-free individuals. During an average of 15.2 years of follow-up, 520 patients were diagnosed with cancer. Lower mtDNA-CN was associated with decreased risk of genital organ cancer (hazard ratio (HR), 0.84), and shorter telomere length was associated with increased risk of urinary system cancer (HR, 1.79). Furthermore, mtDNA-CN was inversely associated with all-cause (HR, 1.20) and cancer-specific mortality (HR, 1.21) when considering all cancer types. Surprisingly, shorter telomere length was associated with decreased risk of cancer-specific mortality when considering all cancer types (HR, 0.85). Finally, lower mtDNA-CN and shorter telomere length were associated with increased risk of both all-cause and cancer-specific mortality in genital organ cancer patients. In this study population, we found that mtDNA-CN and telomere length were significantly associated with prevalent and incident cancer and cancer mortality. However, these associations were cancer type specific and need further investigation.

Keywords: mitochondrial DNA copy number; relative telomere length; cancer types; prevalent cancer; cancer risk; mortality

check for updates

Citation: Li, Y.; Sundquist, K.; Wang, X.; Zhang, N.; Hedelius, A.; Sundquist, J.; Memon, A.A. Association of Mitochondrial DNA Copy Number and Telomere Length with Prevalent and Incident Cancer and Cancer Mortality in Women: A Prospective Swedish Population-Based Study. *Cancers* **2021**, *13*, 3842. https://doi.org/10.3390/cancers13153842

Academic Editor: Izumi Horikawa

Received: 1 July 2021
Accepted: 27 July 2021
Published: 30 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Cancer is expected to rank as the leading cause of death (age < 70) worldwide and is the single most important public health problem that lacks a global solution [1]. Studies from epidemiological profiles of cancer have shown that different exposures to risk factors, lifestyles, economic settings and access to care or screening programs, for a person who develops cancer, may vary heterogeneously [2–4]. Nevertheless, morbidity and mortality caused by cancer in every world region pose a huge threat to global development and lay a tremendous burden on our society.

The dysfunction of mitochondria is one of the hallmarks of cancer. Mitochondria have their own genome (mtDNA, 16,596 base pair) and, according to the energy demands, their copy numbers range from a few hundred to more than 10,000 in a cell type- and origin-specific manner [5,6]. MtDNA is circular double-stranded DNA, located in the mitochondrial inner membrane close to the site where excessive reactive oxygen species (ROS) are routinely generated, and it is prone to be injured by oxidative attack [7–9]. The damaged mtDNA molecules are primarily resolved via robust base excision repair. However, unlike nuclear DNA, mtDNA with double-strand breaks (DSBs) is degraded rapidly instead of being repaired, leading to a significant decrease in mtDNA copy number (mtDNA-CN) [10]. Non-cleaved mtDNA is subsequently replicated by a mitochondrial replisome comprising DNA polymerase gamma (Pol γ), twinkle helicase and single-strand binding protein SSBP1 as a feedback mechanism to compensate for the metabolic defects in impaired mitochondria [11,12]. Thus, mtDNA-CN is a relative measurement that reflects mitochondrial pathologies and it is prone to alteration under various energy requirements and physiological and environmental conditions [13,14]. Nevertheless, erroneous replication and repair can contribute to accumulating mtDNA mutations, leading to mitochondrial dysfunction and signaling to the nucleus [12,15]. As an indirect biomarker for mitochondrial function, mtDNA-CN has been widely associated with many diseases, including cancer [16], aging [17–19], depression [20,21], cardiovascular disease [22,23], type 2 diabetes [24,25], liver disease [26,27], chronic kidney disease [28,29] and neurodegenerative disease [30]. However, current studies on the mtDNA-CN in cancer have reported mixed results; most were based on a case–control design and were inconsistent for various types of cancers.

Telomeres are the nucleoprotein complexes crucial in preserving chromosomal stability and integrity; their length ranges from 5 to 15 kb in humans and varies among tissues [31]. Telomerase is the enzyme responsible for maintaining telomere length and is silenced in normal somatic cells. In the absence of maintenance mechanisms, telomeres undergo shortening with cell division in most human tissues, reflecting organism aging at the cellular level influenced by oxidative stress [32–35]. Short telomeres eventually trigger cellular senescence and a DNA damage signal where cells will stay in a quiescent state for years and secrete factors that influence aging-related diseases rather than undergo apoptosis, which was suggested as a tumor suppressor mechanism for humans [36]. However, the abnormal or extreme shortening of telomere length may cause chromosomal degradation and contribute to malignant cell transformation, which is associated with a higher risk of multiple human diseases, including cancer [37,38]. Telomere shortening has a dual role in carcinogenesis. It promotes the initiation of cancer by inducing chromosomal instability, while telomere length maintenance characterized by telomerase expression is required for cancer cell proliferation and tumor growth [39]. Similar to mtDNA-CN, the reports on telomere length as a biomarker for cancer risk are contradictory.

MtDNA and telomere length are highly variable across cell types but maintained within a constant range according to the specific tissue, therefore, mtDNA-CN and telomere length measured in peripheral blood are considered a surrogate for the measurement of personal health outcomes. Both mitochondria and telomeres serve as critical regulators of the aging process, and their structures are easily damaged by ROS and systemic inflammation; they also play important roles in tumorigenesis [40–42]. The conclusions drawn from previous studies showed conflicting results on the associations between mtDNA-CN

or telomere length and risk of cancers. Possible explanations could be sample collection, sample selection, study design and measurement errors. Few prospective studies have been performed on telomere length and cancer risk. However, to the best of our knowledge, no prospective study has been conducted on mtDNA-CN and all cancer incidence. Furthermore, the most popular techniques for quantification of mtDNA-CN and telomere length are PCR based, which in most cases provide relative measurements. Moreover, there is no study available with a focus on population-based studies systematically analyzing the association between mtDNA-CN, telomere length and the prevalence, incidence and mortality of all cancer types. We aimed to comprehensively explore this possibility in a large cohort of middle-aged Swedish women with precisely quantified mtDNA-CN and telomere length from our well-optimized droplet digital (dd) PCR and quantitative real-time (qRT) PCR methods, respectively. We hypothesized that mtDNA-CN and telomere length are potential biomarkers for the identification of prevalent cancers as well as for the prediction of incident cancers.

2. Materials and Methods

2.1. Study Population

The present study was conducted based on Women's Health in Lund Area (WHILA), a prospective population-based cohort that started in 1995. All women, aged 50–59 years (born between 1935 and 1945) and living in Scania in southern Sweden, were invited to participate in a health survey. From December 1995 to February 2000, a total of 6917 women (out of 10,766, the total population of women in the five southern municipalities in 1995) underwent a physical examination and answered a questionnaire. There was no financial reimbursement for participation. After providing written consent, the participants were given up to two hours to answer the questionnaire. The questionnaire that was distributed to all participants has been described previously [43]. If they had any uncertainties, they could ask an experienced research nurse for assistance. Participants were followed from the day of screening until death, or if no event occurred, until 31 May 2015. However, the blood samples for DNA extraction were collected midway through this study (from October 1997) and therefore 3225 participants were included in the present study.

2.2. Outcome Measurement

Information about cancer incidence and mortality was obtained from the Swedish Cancer Registry and Death Registry and information on prevalent cancer was obtained from self-reported data from questionnaires. Among the participants included in the study, 187 individuals were diagnosed with cancer at baseline (prevalent cancer) and 3038 individuals were cancer-free at baseline. We followed the cancer-free women from the day of screening until (1) cancer diagnosis; (2) death; (3) ending date of this study (31 May 2015). Individuals' diagnoses of cancer were then identified and followed until death from any cause (overall mortality) and from cancer (cancer mortality) and/or till the end of the study period, whichever came first.

The following cancer outcomes were classified according to the WHO's International Classification of Diseases (revision 10) as (a) breast cancer; (b) digestive system cancer (liver cancer, pancreatic cancer, gastric cancer, small intestine cancer, rectum cancer, colon cancer and oral cancer); (c) respiratory system cancer (lung cancer); (d) genital organ cancer (ovary cancer, cervix cancer, uterus cancer and corpus cancer); (e) urinary system cancer (kidney cancer and urethral cancer); (f) hematological cancer (myeloma, leukemia and non-Hodgkin's lymphoma); (g) nervous system cancer; (h) melanoma and other malignant neoplasms of the skin; (i) endocrine gland cancer (thyroid cancer).

2.3. Extraction of DNA

Peripheral blood samples were collected in ethylenediaminetetraacetic acid (EDTA) tubes. Total genomic DNA was extracted using a QiAamp96 DNA Blood (Qiagen, Inc., Hilden, Germany) from a 200 µL blood sample according to the manufacturer's instructions.

The concentrations and purities of isolated DNA samples were spectrometrically analyzed and frozen at −20 °C for further usage.

2.4. Quantification of Relative Telomere Length

Genomic DNA extracted from blood was quantified by a Nanodrop (ND-2000, Thermo Scientific, Waltham, MA, USA) and then normalized to 5 ng/uL in TE buffer that contained Escherichia coli DNA. The DNA was heated at 95 °C for 30 min, then followed by 1 min on ice, spun down briefly at 1000× *g* at 4 °C and kept at 4 °C. Telomere length was measured by real-time PCR based on a previous report by Cawthon [44] and modified by our group.

The copy number of telomeric repeats was compared to a single copy gene (β-hemoglobin, HBG) to normalize the quantity of the input DNA. The telomere to HBG (T/S) ratio represents the average relative length of the telomeres. Detailed methods have been described previously [45]. Briefly, 20 ng DNA from samples and 7 references (from Jurkat cell line) were pooled in triplicate in 384-well plates, qPCR was performed separately for telomeres and HBG and negative controls were included. As for the measurement of telomere length, a standard curve from reference DNA was generated (Bio-Rad CFX Manager software v. 2.0.) and used in each assay plate. Telomere and HBG concentrations were calculated according to the standard curve. All standard curves for both telomere and HBG had correlation coefficients of $R^2 > 0.99$. The PCR efficiencies for each reaction were higher than 93%. The inter- and intracoefficients of variation (CV) for the T/S ratios were 6.2% and 3%, respectively.

2.5. Quantification of MtDNA Copy Number

Droplet digital PCR (ddPCR) was used to quantify the absolute copy number of nuclear DNA (nDNA) and mtDNA. The mtDNA/nDNA content was assessed using specific primers designed to target the mitochondrial MT-ND1 (assay ID: dHsaCPE5029120) gene and nuclear EIF2C1 (assay ID: dHsaCP1000002) gene. Probes targeting nDNA were attached to a HEX fluorophore whereas mtDNA was attached to FAM and had an Iowa Black® FQ quencher on all probes. All primer and probes were obtained from Bio-Rad (Hercules, CA, USA). Quality control for every step of our well-optimized ddPCR method was stringent, as described previously [46]. Briefly, 1ng DNA from samples, including positive and negative controls, was separately pooled in a 20 uL multiplex reaction containing primers (900 nM), probes (250 nM), ddPCR Supermix for probes (no UTP, 2X) and 5U/reaction restriction enzyme (HindIII). The plate with reactions was sealed and incubated at room temperature for 20 min to allow restriction enzyme digestion and then loaded into the automated droplet generator to generate droplets, followed by end-point PCR. The after-PCR plate was kept overnight at 4 °C to maximize the droplet recovery. The plate was finally read on the droplet reader, and data were collected and analyzed using QuantaSoft™ Software to calculate the numbers of positive and negative droplets in each sample. The fraction of positive droplets was then fitted to a Poisson distribution to determine the absolute copy number in units of copies/µL. The inter- and intra-CVs for absolute quantification of mtDNA-CN were 4.2% and 3.1%, respectively.

2.6. Assessment of Covariates

We collected information on potential confounding factors at baseline through the health survey, including age at screening, body mass index (BMI), education (1–9, 10–11, ≥12 years of schooling), alcohol habits (no consumption, <12 g/day, ≥12 g/day) and smoking habits as non-smokers, past smokers (≥1 pack year, stopped smoking ≥1 month prior to the study) and current smokers (≥1 pack year). Physical activity at home was defined according to the questionnaire and the participants with a score of 1–3 were categorized as low activity at home: 1 = hardly do anything at all, 2 = mostly sedentary, 3 = light physical exertion. High activity at home was categorized with a score of 4–6: 4 = strenuous exercise 1–2 h/week, 5 = strenuous exercise at least 3 h/week, 6 = hard regular exercise. Physical activity at work was categorized as low, moderate and high. Information

on comorbidity was collected from both baseline self-reported questionnaires and the Swedish health registries concerning diabetes (including type 1 and type 2 diabetes as yes/no), cardiovascular disease (CVD, including stroke, coronary heart disease, abdominal aortic aneurysm as yes/no) and hypertension (yes/no).

2.7. Statistics

A Pearson chi-square test was used to compare categorical variables (education level, smoking habits, alcohol consumption, activity at work, activity at home, diabetes, CVD, hypertension) and continuous variables (age at screening, BMI) were compared using Student's *t*-tests. Linear regression analysis was performed to evaluate the association between prevalent cancer (yes/no) and mtDNA-CN or telomere length at baseline. We further produced a Cox proportional hazards model to explore the association between mtDNA-CN, telomere length and cancer incidence in 3038 cancer-free individuals. Subjects were dichotomized into high and low mtDNA or long and short telomere length groups according to the median based on the distribution of mtDNA-CN or telomere length. The high or long group served as the reference group in the analyses. Hazard ratios (HRs) and 95% confidence intervals (95% CIs) were calculated to evaluate the association between mtDNA-CN, telomere length and cancer risk. We further examined the association between mtDNA-CN, telomere length and all-cause mortality, as well as cancer-specific mortality in 520 cancer patients. Competing risk models were created while analyzing cancer-specific mortality. Deaths from other causes were considered as competing risks. Kaplan–Meier survival curves were calculated to evaluate the association between mtDNA-CN and telomere length and cancer mortality. To control for potential confounders, the following variables were included in the multivariate regression model: age, BMI, education level, smoking habits, alcohol consumption, activity at work, activity at home, diabetes, CVD and hypertension. All statistical analyses were carried out in SPSS software version 23 (IBM, Armonk, NY, USA) and SAS version 9.4.

3. Results

Of the 3225 participants who had their blood samples collected at baseline and were included in this study, 187 (5.8%) were reported as having prevalent cancer and 3038 women without cancer were followed for incident cancer. During an average 15.2 years of follow-up, 520 of 3038 participants (17.1%) developed cancer and, among them, 138 died during the follow-up (Figure 1).

3.1. Population Characteristic of Prevalent Cancer and No Cancer at Baseline

Table 1 shows the characteristics of the study population at baseline. Compared to cancer-free individuals, cases were older and less likely to consume alcohol ($p < 0.05$). Telomere length (mean ± SD) was normally distributed and was shorter in participants with prevalent cancer. No significant differences were observed between cancer-free individuals and prevalent cancer patients in terms of BMI, education level, smoking habits, activity at work, activity at home, diabetes, CVD or hypertension. All of the variables referenced above were considered as potential confounders and were adjusted in the subsequent multivariable analyses.

3.2. Prevalent Cancer and MtDNA-CN/Telomere Length

We performed further crude and adjusted linear regression analysis to investigate the association between prevalent cancer and mtDNA-CN or telomere length. The cancer diagnoses were categorized into cancer types according to ICD codes to determine whether the results applied to site-specific cancers. All cancers in this study were categorized across the nine main cancer types in the following way: breast, digestive system, respiratory system, genital organ, urinary system, hematological tumor, nervous system, skin and endocrine gland cancer (Table S1).

Our results show that prevalent breast cancer was significantly associated with higher mtDNA-CN (adjusted β was 12.39; 95% CI = 4.15, 20.63; $p = 0.003$). An inverse association between prevalent hematological cancer and mtDNA-CN was found, however, it did not reach statistical significance (adjusted β was −25.21; 95% CI = −51.19, 0.77; $p = 0.057$) (Table 2).

Furthermore, prevalent cancer was significantly associated with shorter telomere length (crude β was −0.03; 95% CI = −0.05, −0.01; $p = 0.027$). However, this association became non-significant after adjusting for potential confounders (adjusted β was −0.02; 95% CI = −0.04, 0; $p = 0.059$). Stratification of the data according to the cancer types suggested that the associations between telomere length and breast and genital cancers were stronger; however, the results did not reach statistical significance (Table 3).

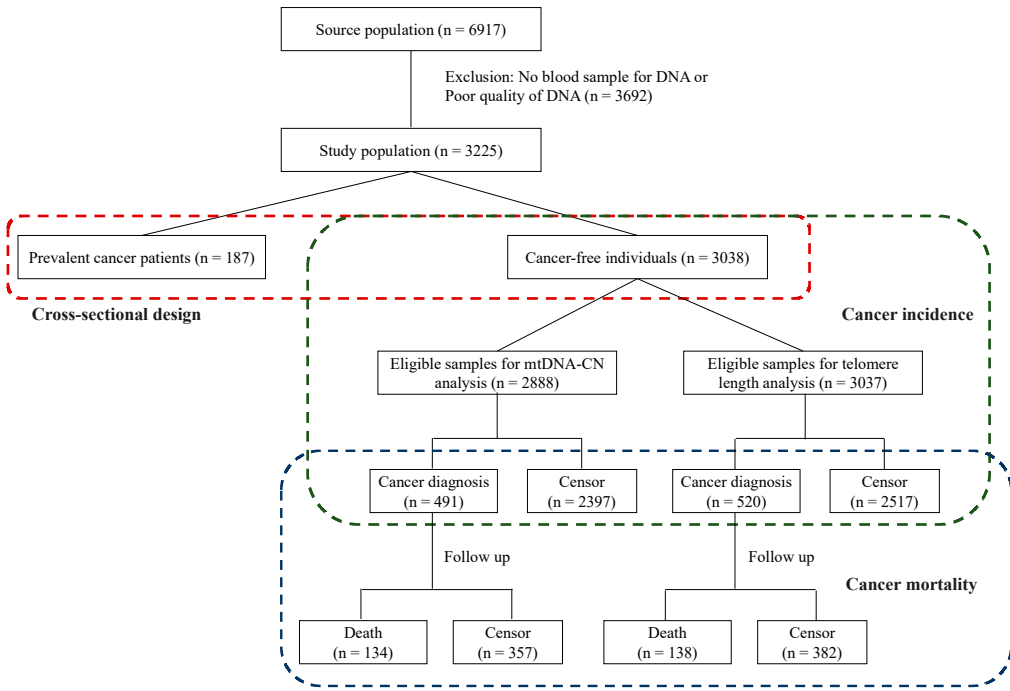


Figure 1. Flow chart of participants in this cohort study.

Table 1. Baseline characteristics of mtDNA-CN and telomere length stratified by prevalent and no cancer.

Characteristics	Prevalent Cancer (n = 187)		No Cancer (n = 3038)		p-Value [a]
	Mean	**SD**	**Mean**	**SD**	
Age	57.6	2.8	57.1	2.9	0.027
BMI	25.7	4.3	25.7	4.1	0.848
mtDNA-CN	111.7	40.8	109.0	35.1	0.385
Telomere length	0.69	0.15	0.72	0.15	0.027
	Number	**%**	**Number**	**%**	
Education level					0.755
0–9	112	57.1	1736	59.9	
10–11	25	14.0	424	13.4	
≥12	50	28.9	878	26.7	
Smoking habit					0.585
Non-smokers	151	79.9	2422	80.7	
Past smokers	6	1.8	55	2.7	
Current smokers	31	18.5	561	16.6	
Alcohol habit					0.048
No consumption	62	33.2	761	25.0	
<12 g/day	105	56.1	1908	62.8	
≥12 g/day	20	10.7	369	12.1	
Activity at work					0.866
Low	51	27.3	866	28.5	
Moderate	88	47.1	1441	47.4	
High	48	25.7	731	24.1	
Activity at home					0.174
Low	119	63.6	1780	58.6	
High	68	36.4	1258	41.4	
Comorbidity					
Diabetes	34	18.2	416	13.7	0.086
CVD	39	20.9	547	18.0	0.327
Hypertension	79	42.2	1266	41.7	0.877

CVD indicates cardiovascular disease. [a] Student's *t*-tests were performed for continuous variables. Chi-square tests were performed for categorical variables.

Table 2. Linear regression models examining association between prevalent cancer and mtDNA-CN.

Characteristics	No. of Cases	Crude Univariate		Adjusted Univariate [a]	
		B (95% CI)	***p*-Value**	**β (95% CI)**	***p*-Value**
Prevalent cancer	174	2.75 (−2.67, 8.18)	0.320	4.18 (−1.18, 9.55)	0.059
Breast cancer	71	10.70 (2.36, 19.03)	0.012	12.39 (4.15, 20.63)	0.003
Digestive system	11	−6.53 (−27.52, 14.46)	0.351	−6.54 (−27.27, 14.20)	0.537
Respiratory system	1	3.86 (−65.64, 73.36)	0.913	1.27 (−67.36, 69.90)	0.971
Genital organs	30	−3.48 (−16.23, 9.27)	0.593	−0.48 (−13.09, 12.13)	0.941
Urinary system	5	−24.98 (−56.07, 6.11)	0.115	−20.77 (−51.47, 9.94)	0.185
Hematological cancer	7	−25.23 (−51.51, 1.05)	0.060	−25.21 (−51.19, 0.77)	0.057
Nervous system	5	−10.76 (−41.86, 20.34)	0.498	−8.80 (−39.53, 21.93)	0.575
Skin	17	11.27 (−5.62, 28.17)	0.191	9.51 (−7.19, 26.21)	0.264
Endocrine glands	17	−2.57 (−19.47, 14.47)	0.766	−0.88 (−17.58, 15.84)	0.918

[a] Adjusted for age, body mass index (BMI), education level, smoking habits, alcohol consumption, activity at work, activity at home, diabetes, CVD, hypertension.

Table 3. Linear regression models examining association between prevalent cancer and telomere length.

Characteristics	No. of Cases	Crude Univariate		Adjusted Univariate [a]	
		β (95% CI)	*p*-Value	β (95% CI)	*p*-Value
Prevalent cancer	187	−0.03 (−0.05, −0.01)	0.027	−0.02 (−0.04, 0)	0.059
Breast cancer	77	−0.03 (0.06, 0.01)	0.098	−0.03 (−0.06, 0.01)	0.143
Digestive system	11	−0.02 (−0.11, 0.07)	0.697	−0.02 (−0.11, 0.07)	0.683
Respiratory system	1	−0.08 (−0.38, 0.22)	0.592	−0.08 (−0.38, 0.22)	0.597
Genital organs	35	−0.05 (−0.10, 0)	0.069	−0.04 (−0.09, 0.01)	0.119
Urinary system	5	−0.05 (−0.18, 0.09)	0.472	−0.03 (−0.17, 0.10)	0.635
Hematological cancer	7	0.08 (−0.04, 0.19)	0.175	0.08 (−0.04, 0.19)	0.191
Nervous system	5	−0.08 (−0.21, 0.05)	0.219	−0.08 (−0.21, 0.06)	0.256
Skin	18	−0.02 (−0.09, 0.06)	0.687	−0.02 (−0.09, 0.05)	0.628
Endocrine glands	17	−0.01 (−0.09, 0.06)	0.704	−0.01 (−0.08, 0.07)	0.834

[a] Adjusted for age, BMI, education level, smoking habits, alcohol consumption, activity at work, activity at home, diabetes, CVD, hypertension.

3.3. Cancer Incidences and MtDNA-CN/Telomere Length

Baseline characteristics of participants with no cancer at baseline are shown in Table S2. MtDNA-CN and telomere length were normally distributed. A significant decrease in mtDNA-CN and telomere length was seen with age ($p < 0.001$, Figures S1 and S2). For mtDNA-CN, further associations were observed for the following variables: education level, smoking habits, alcohol habits, activity at work, diabetes, CVD. Baseline telomere length was shorter in participants with higher BMI and less physical activity (Table S2).

During an average of 15.2 years of follow-up, we identified 520 patients with a cancer diagnosis. To determine if the level of mtDNA-CN or telomere length was associated with cancer risk, single-factor Cox regression analyses were conducted (Table 4). Participants with a lower level of mtDNA-CN at baseline had a lower risk of having genital organ cancer during follow-up, and the hazard ratio (HR) per one standard deviation (SD) decrease in mtDNA-CN for incident genital organ cancer was 0.84 (95% CI = 0.72, 0.98). Individuals with lower mtDNA-CN had increased risks of developing urinary system cancer (adjusted HR 8.2, 95% CI, 1.06–63.2) and hematological cancer (adjusted HR 1.97, 95% CI, 1.02–3.81). No other cancer type was significantly associated with mtDNA-CN. For a 1 SD decrease in the telomere length, the risk for incident urinary system cancer increased 1.79 times (adjusted HR 1.79, 95% CI = 1.05, 3.07). The results showed a similar trend when dichotomizing the mtDNA-CN (low, ≤111 copies/μL; high, >111 copies/μL) and telomere length (short, ≤0.721965; long, >0.721965) according to the median into two groups. Furthermore, the interactions for mtDNA-CN and telomere length for urinary system cancer and hematological cancer were statistically significant.

3.4. Mortality and MtDNA-CN/Telomere Length

During the follow-up of 520 cancer patients, a total of 138 participants died (all-cause mortality), and we also investigated the association between cancer mortality and mtDNA-CN or telomere length. The Kaplan–Meier plots are presented in Supplementary Figures S3 and S4.

We found that lower mtDNA-CN at baseline was associated with increased all-cause mortality (multivariable HR per 1 SD decrease, 1.20; 95% CI = 1.01, 1.42) as well as cancer-specific mortality when considering all cancer types (multivariable HR per 1 SD decrease, 1.21; 95% CI = 1.01, 1.45). Stratification of data, according to cancer type, showed an association between mtDNA-CN and all-cause mortality and cancer-specific mortality in genital cancer patients. The risk for all-cause mortality increased 2.15 times (adjusted HR 2.15, 95% CI = 1.04, 4.44) and cancer-specific mortality increased 2.42 times (adjusted HR 2.42, 95% CI = 1.03, 5.70) for a 1 SD decrease in mtDNA-CN after adjusting for potential confounders. We also dichotomized mtDNA-CN levels according to the median and our results showed that compared with participants in the higher mtDNA-CN group, the multivariable HR for mortality from all causes in genital cancer patients was 8.06 (95% CI = 1.75, 37.2) and mortality from specific genital cancer was 5.59 (95% CI = 1.61, 19.4) in the lower level mtDNA-CN group (Table 5).

Table 4. Hazard ratios and 95% confidence intervals of cancer incidence associated with mtDNA-CN (n = 2888) and relative telomere length (n = 3037).

Characteristics	MtDNA-CN		HR (95% CI) per 1-SD Decrease in mtDNA-CN	Relative Telomere Length		HR (95% CI) per 1-SD Decrease in Telomere Length	p for Interaction [c]
	High (n = 1406)	Low (n = 1482)		Long (n = 1473)	Short (n = 1564)		
All cancer							
No. of cancer diagnoses	242	249		256	264		
Person-years of follow-up	21,700	22,238		23,394	22,626		
IR, per 1000 person-years	11.15	11.19		10.94	11.67		
Crude HR (95% CI)	1 (Ref)	1.04 (0.87–1.24)	1.01 (0.92–1.10)	1 (Ref)	0.97 (0.82–1.16)	1.02 (0.94–1.11)	
Adjusted HR (95% CI) [a]	1 (Ref)	0.99 (0.83–1.19)	0.99 (0.90–1.08)	1 (Ref)	0.93 (0.78–1.10)	1.00 (0.92–1.09)	0.607
Cancer types (adjusted HR and 95% CI [a])							
Breast cancer	1 (Ref)	1.05 (0.78–1.42)	0.95 (0.84–1.08)	1 (Ref)	1.03 (0.77–1.38)	0.98 (0.84–1.14)	0.197
Digestive system	1 (Ref)	0.83 (0.54–1.29)	1.03 (0.84–1.26)	1 (Ref)	1.11 (0.71–1.72)	1.05 (0.88–1.25)	0.225
Respiratory system	1 (Ref)	0.91 (0.42–2.02)	1.18 (0.88–1.60)	1 (Ref)	0.58 (0.29–1.16)	0.83 (0.63–1.10)	0.192
Genital organs	1 (Ref)	0.60 (0.33–1.10)	0.84 (0.72–0.98) [b]	1 (Ref)	1.16 (0.66–2.05)	1.13 (0.87–1.47)	0.207
Urinary system	1 (Ref)	8.20 (1.06–63.2) [b]	1.08 (0.72–1.63)	1 (Ref)	1.35 (0.45–4.08)	1.79 (1.05–3.07) [b]	<0.001
Hematological cancer	1 (Ref)	1.97 (1.02–3.81) [b]	1.11 (0.85–1.46)	1 (Ref)	0.81 (0.43–1.54)	0.97 (0.71–1.32)	0.007
Nervous system	1 (Ref)	1.62 (0.49–5.31)	1.07 (0.66–1.72)	1 (Ref)	0.68 (0.22–2.07)	0.74 (0.45–1.22)	0.314
Skin	1 (Ref)	0.82 (0.49–1.39)	1.05 (0.82–1.34)	1 (Ref)	0.81 (0.50–1.32)	1.11 (0.88–1.38)	0.171
Endocrine glands	1 (Ref)	0.79 (0.29–2.13)	1.01 (0.63–1.63)	1 (Ref)	0.67 (0.27–1.69)	0.76 (0.49–1.16)	0.331

[a] Adjusted for age, BMI, education level, smoking habits, alcohol consumption, activity at work, activity at home, diabetes, CVD, hypertension. [b] $p < 0.05$. [c] Interactions were calculated by inclusion of interaction terms. IR, incidence rate.

Table 5. Hazard ratios and 95% confidence intervals of mortality associated with mtDNA-CN (n = 491) and relative telomere length (n = 520) among cancer patients.

Characteristics	MtDNA-CN		HR (95% CI) per 1-SD Decrease in mtDNA-CN	Relative Telomere Length		HR (95% CI) per 1-SD Decrease in Telomere Length	p for Interaction [c]
	High	Low		Long	Short		
All cancer							
No. of cancer patients	242	249		256	264		
No. of all-cause deaths	62	72		74	64		
No. of cancer-specific deaths	57	66		69	58		
Person-years of follow-up	1652	1578		1666	1787		
All-cause mortality rate, per 100 person-years	3.75	4.56		4.44	3.58		
Cancer specific mortality rate, per 100 person-years	3.45	4.18		4.14	3.25		
Adjusted all-cause mortality HR (95% CI)	1 (Ref)	1.14 (0.80–1.62)	1.20 (1.01–1.42) [b]	1 (Ref)	0.78 (0.55–1.10)	0.86 (0.72–1.02)	0.090
Adjusted cancer-specific mortality HR (95% CI) [a]	1 (Ref)	1.15 (0.80–1.66)	1.21 (1.01–1.45) [b]	1 (Ref)	0.75 (0.53–1.07)	0.85 (0.71–1.00) [b]	0.117
Cancer types (adjusted all-cause mortality HR and 95% CI) [a]							
Breast cancer	1 (Ref)	1.42 (0.58–3.51)	1.16 (0.77–1.74)	1 (Ref)	0.61 (0.25–1.49)	0.73 (0.46–1.14)	0.952
Digestive system	1 (Ref)	1.08 (0.46–2.51)	1.53 (1.02–2.28)	1 (Ref)	1.36 (0.63–2.96)	0.86 (0.54–1.38)	0.959
Respiratory system	1 (Ref)	0.27 (0.06–1.25)	0.48 (0.18–1.24)	1 (Ref)	0.54 (0.14–2.20)	0.97 (0.50–1.89)	0.208
Genital organs	1 (Ref)	8.06 (1.75–37.2) [b]	2.15 (1.04–4.44) [b]	1 (Ref)	3.73 (0.86–16.2)	2.12 (1.00–4.52) [b]	0.503
Urinary system	1 (Ref)	-	-	1 (Ref)	-	-	
Hematological cancer	1 (Ref)	0.05 (0.00–74.4)	-	1 (Ref)	0.24 (0.30–1.89)	-	
Nervous system	1 (Ref)	-	-	1 (Ref)	-	-	
Skin	1 (Ref)	-	-	1 (Ref)	-	-	
Endocrine glands	1 (Ref)	-	-	1 (Ref)	-	-	

Table 5. *Cont.*

Characteristics	MtDNA-CN		HR (95% CI) per 1-SD Decrease in mtDNA-CN	Relative Telomere Length		HR (95% CI) per 1-SD Decrease in Telomere Length	*p* for Interaction [c]
	High	Low		Long	Short		
Cancer types (adjusted cancer-specific mortality HR and 95% CI) [a]							
Breast cancer	1 (Ref)	1.71 (0.67–4.35)	1.23 (0.86–1.77)	1 (Ref)	0.34 (0.11–1.04)	0.56 (0.34–0.95) [b]	0.829
Digestive system	1 (Ref)	0.89 (0.34–2.39)	1.42 (0.90–2.25)	1 (Ref)	1.06 (0.46–2.44)	0.80 (0.48–1.34)	0.609
Respiratory system	1 (Ref)	0.34 (0.08–1.44)	0.48 (0.21–1.08)	1 (Ref)	0.83 (0.32–2.15)	1.19 (0.72–1.96)	0.421
Genital organs	1 (Ref)	5.59 (1.61–19.4) [b]	2.42 (1.03–5.70) [b]	1 (Ref)	4.56 (1.30–16.0)	1.98 (1.10–3.53) [b]	0.167
Urinary system	1 (Ref)	-	-	1 (Ref)	-	-	
Hematological cancer	1 (Ref)	-	-	1 (Ref)	-	-	
Nervous system	1 (Ref)	-	-	1 (Ref)	-	-	
Skin	1 (Ref)	-	-	1 (Ref)	-	-	
Endocrine glands	1 (Ref)	-	-	1 (Ref)	-	-	

[a] Adjusted for age, BMI, education level, smoking habits, alcohol consumption, activity at work, activity at home, diabetes, CVD, hypertension. [b] $p < 0.05$. [c] Interactions were calculated by inclusion of interaction terms. IR, incidence rate. -, not enough cases to conduct the analysis.

The HR per 1 SD decrease in telomere length for cancer-specific mortality was 0.85 (95% CI, 0.21–1.00, multivariable model) when considering all cancer types. Similar to mtDNA-CN, shorter telomere length at baseline was associated with increased risk of all-cause mortality (adjusted HR per 1 SD decrease was 2.23; 95% CI = 1.00, 4.52) and cancer-specific mortality (adjusted HR per 1 SD decrease was 1.98; 95% CI = 1.10, 3.53) in genital cancer patients. Of note, both a lower level of mtDNA-CN and shorter telomere length were preferentially associated with increased mortality in patients with genital cancer such as ovary, cervix, uterus and corpus cancer.

4. Discussion

To the best of our knowledge, this prospective cohort study is the first population-based study to comprehensively explore the association between mtDNA-CN and telomere length and cancer prevalence and incidence, as well as cancer mortality, among middle-aged women. Our results show that both mtDNA-CN and telomere length are associated with the prevalence as well as with future risk of cancer but in a cancer-specific manner. Our results also show that mtDNA-CN was inversely associated with all-cause mortality and cancer-specific mortality when considering all cancer types. Finally, shorter telomere length was associated with a lower risk of cancer-specific mortality in all cancer types and breast cancer. However, in genital cancer, lower mtDNA-CN and shorter telomere length were associated with increased risk of all-cause mortality and cancer-specific mortality.

4.1. Comparison with Previous Studies

4.1.1. MtDNA-CN, Relative Telomere Length and Prevalent Cancer

Although we found that there was no significant association between mtDNA-CN and overall prevalent cancer, we demonstrated that patients with prevalent breast cancer had higher mtDNA-CN compared with cancer-free individuals. Consistent with our result, a meta-analysis including 21 prospective studies and 17 retrospective case–control studies also suggested no significant association between mtDNA-CN and overall prevalent cancer [47]. Together, these results suggest that the association between higher mtDNA-CN and prevalent cancer may be study population and cancer type specific and this could be one of the reasons for the conflicting results published to date.

We found that prevalent cancer was associated with shorter telomere length, but this association decreased after adjusting for potential confounders. In agreement with our result, a meta-analysis of 46 retrospective observational studies also demonstrated a borderline significant relationship between telomere length and overall prevalent cancer [48].

4.1.2. Baseline Levels of MtDNA-CN, Relative Telomere Length and Cancer Incidence

Our results demonstrated that a lower level of baseline mtDNA-CN was associated with a lower future risk of genital organ cancer, urinary system cancer and hematological cancer. Thus far, few prospective studies have been performed to investigate the association between baseline mtDNA-CN and future cancer risk. Consistent with our results, a nested case–control study observed a positive association between mtDNA-CN and risk in renal cells [49]. Furthermore, two prospective studies also supported the hypothesis that higher mtDNA-CN was associated with increased risk of chronic lymphocytic leukemia/small lymphocytic lymphoma [50] and non-Hodgkin's lymphoma [51].

We observed that women with shorter telomere length had a higher risk of urinary system cancer. However, the result from another prospective study did not support a significant association between leukocyte telomere length and future risk of renal cell carcinoma [52].

4.1.3. Baseline MtDNA-CN, Relative Telomere Length and Mortality in Cancer Patients

Our results show that lower baseline mtDNA-CN was associated with increased all-cause mortality as well as cancer-specific mortality in all cancer types, which is consistent

with the result of the previous meta-analysis [53]. When categorized according to cancer type, lower baseline mtDNA-CN was associated with genital cancer mortality.

We observed heterogeneous associations between telomere length and mortality in different cancer types. Shorter baseline telomere length was associated with increased all-cause mortality as well as cancer-specific mortality in genital organ cancer, but decreased cancer-specific mortality in breast cancer. These inconsistent results across cancer types may reflect different carcinogenic mechanisms conferred by specific telomeres in specific cancer types. A previous systematic review suggested that shorter telomere length was associated with poorer outcomes, which supported our result [54]. Shanta et al. reported a significant association between shorter telomere length and poorer overall survival and progression-free survival in patients with ovary cancer and cervical cancer [55]. However, another study from Kotsopoulos et al. did not support a significant relationship in ovary cancer patients [56].

Smoking is known to be significantly associated with the risk of several cancer types such as cancer in the respiratory system, digestive system and urinary system [57]. Smoking was also inversely associated with both mtDNA-CN and telomere length [9,58]. We further investigated the association between cancer incidence and/or mortality and mtDNA-CN and/or telomere length stratified by smoking status. Our result showed that the risk of all-cause mortality increased 2.63 times (95% CI = 1.19, 5.83) and cancer-specific mortality 2.59 times (95% CI = 1.17, 5.42) for current or past smokers with low mtDNA-CN levels (Table S3).

4.2. Potential Biological Mechanisms

Mitochondria are essential organelles that generate energy in the form of ATP through respiration and oxidative phosphorylation (OXPHOS), produce ROS and initiate and execute apoptosis. In cancers, the malfunctioning mitochondria shift metabolism from OXPHOS to aerobic glycolysis, which has been suggested as a hallmark of carcinogenesis [59]. Mitochondrial dysfunction links to a decrease in apoptosis, an elevated level of ROS and the activation of the hypoxia-like pathway, which also affects nuclear gene expression and methylation [60,61]. MtDNA-CN, as a proxy for mitochondria function, has been shown to differ between cancer tissues and corresponding normal tissues for a number of cancer types, and its alterations in cancer appear to be tissue and tumor stage specific [62]. In addition, average mtDNA-CN levels in blood decrease after the age of 50 in healthy people [63]. However, little is known about the mechanisms that lead to the alteration in mtDNA-CN and the factors involved in the tissue-specific changes in cancers. Extensive genetic studies offer evidence that polymorphic mutations are significantly associated with mtDNA-CN levels and they seem to be context specific [64–66]. In addition to genetic factors, a few studies also showed that different exposures to various chemicals, risk factors, lifestyles, economic settings and health care systems significantly influence mtDNA-CN [67–70]. Thus, the changes in mtDNA-CN might directly depend on the type of mutations in nuclear DNA or mtDNA and/or be an adaptive response towards the effect of the mutations in order to gain a growth advantage for certain types of tumors [71].

Telomeres are specialized structures that protect the ends of chromosomes from fusion and DNA damage. Telomere length—a complex hereditary trait—seems to be a mitotic clock of the lifespan of the cells; its maintenance has been widely studied but is not well understood [72]. The telomerase enzyme plays a dominant role in maintaining and regulating telomere length, and is upregulated in tumors compared with normal tissue counterparts in over 90% of cancers. A subset of tumors employ a telomerase-independent, homologous recombination-based mechanism called alternative lengthening of telomeres (ALT) to elongate telomere length [73]. In cancers, a paradox about telomere length exists; individuals with long telomeres have a higher risk for the majority of cancers while cancerous tissues have short telomeres. Given that aging is the major risk factor for cancers, telomeres in somatic cells are typically shorter in older populations [74]. Short telomere length combined with other oncogenic changes might impair immune surveillance and lead

to carcinogenesis [75]. Tumor cells that undergo oncogenic changes continue to divide and bypass the senescence, and this stage is accomplished by either upregulation or reactivation of telomerase expression, or by acquiring rarer ALT mechanisms to maintain these very short telomeres to achieve cell immortality [72]. GWASs and other studies conducted on different populations reported the identification of 18 multiple SNPs and rare variants that were associated with telomere length [76,77]. Some studies have also shown that carcinogen exposure, oxidative stress, inflammation, lifestyle and physiological stress were associated with telomere dynamics [76].

Mitochondrial DNA and telomeres have been implicated in the aging process for a long time. Growing evidence shows that telomere attrition regulates mitochondrial biogenesis and function through the PARP1-NAD^+-SIRT1, ATM/R-P53-PGC1α/β and ATM-AKT-mTOR-PGC1β pathways, eventually resulting in mitochondrial dysfunction and increased ROS generation [78]. Beyond aging, studies have also revealed the importance of the telomere–p53–mitochondrial *axis* for cancer [79]. Therefore, further research is necessary to elucidate the biological mechanism underlying the telomere and mitochondrion connection.

4.3. Clinical Relevance

Our findings suggest that potential clinical applications of mtDNA-CN or telomere length as tools for risk prediction of different cancers and mtDNA-CN might be used as a prognostic biomarker of malignancy. For example, based on our results, for genital organ cancers, which are hormone-associated cancers, postmenopause with a lower level of baseline mtDNA-CN or shorter telomere length will be suggested as an indicator of poor health status, and therefore we can identify individuals for adjustment of lifestyle or for hormone replacement therapy. However, the conflicting associations between mtDNA-CN, telomere length and risk of cancer suggest that the application of these biomarkers to the general population may be premature at this stage.

4.4. Strengths and Limitations

As a whole, our study has important merits. This is a population-based study with a cohort followed prospectively for up to 20 years with a large population size. Second, we performed the analysis only on middle-aged women and therefore it is not confounded by variations in age and sex. Third, methodological bias is one of the main factors for conflicting results published to date, but here, we used our well-optimized methods for measurement of mtDNA-CN and telomere length to make the findings more consistent and reliable. Compared to real-time PCR, our well-established ddPCR method does not require external standards, has greater precision and improved reproducibility to provide a rigorous quantification of the absolute mtDNA-CN [46]. Fourth, cancer cases were defined by a questionnaire and the Swedish Cancer Register, so we had complete information on cancer diagnosis and death during long-term follow-up.

Nevertheless, there are a few limitations to our study. First, for participants with prevalent cancer, we do not have information on whether they underwent any chemotherapy or radiation therapy when the blood samples were drawn at baseline. Previous studies indicate that cancer treatment alters mtDNA-CN and telomere length [80–82]; thus, we cannot completely rule out the influence of treatment on the changes in mtDNA-CN and telomere length. Second, although our sample size is sufficient for the overall analysis, it is limited to specific cancer types with a small number of cases. The power was also limited for the analysis of prevalent cancer.

5. Conclusions

To the best of our knowledge, this is the first molecular epidemiological study in which we have simultaneously investigated the associations between mtDNA-CN, telomere length and prevalence, incidence and mortality of all cancer types in a large population-based prospective study. Our study strengthens several previous relevant findings and extends our understanding of the link between mtDNA-CN, telomere length and future

risk of several cancer types. Further research is required to validate our results before the application of mtDNA-CN and telomere length as cancer biomarkers.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/cancers13153842/s1, Table S1: Spectrum of cancers in the WHILA study (1995–2015). Table S2: Baseline characteristics of cancer incidence. Table S3: Hazard ratios and 95% confidence intervals of cancer incidence and mortality associated with mtDNA-CN and relative telomere length stratified by smoking status, Figure S1: The coefficient (r) correlation between mtDNA-CN and age was −0.126 ($p < 0.001$), Figure S2: The coefficient (r) correlation between relative telomere length and age was −0.139 ($p < 0.001$), Figure S3: Kaplan-Meier plot for cancer mortality by mtDNA-CN categorized into two groups according to the median, Figure S4: Kaplan-Meier plot for cancer mortality by relative telomere length categorized into two groups according to the median.

Author Contributions: K.S., J.S. and A.A.M. conceived and designed the study; Y.L. and A.H. performed the experiments and Y.L., N.Z. and A.A.M. performed the statistical analysis; K.S., J.S., A.H. and A.A.M. collected the samples and clinical data. Y.L., X.W., N.Z. and A.A.M. participated in data analysis and interpretation; Y.L. wrote the first draft and J.S., X.W., A.H., K.S. and A.A.M. revised the article. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Swedish Research Council and ALF funding from Region Skåne to Jan Sundquist and Kristina Sundquist and by Allmänna Sjukhusets i Malmö Stiftelse för bekämpande av cancer awarded to Ashfaque Memon.

Institutional Review Board Statement: The regional ethical committee at Lund University approved the study (approval nos. 95/174, 2011/494 and 2015/6) according to the guidelines of the Declaration of Helsinki.

Informed Consent Statement: Written informed consent has been obtained from the all the participants to publish this paper.

Data Availability Statement: The data that support the findings of this study are available from the Swedish National Board of Health and Welfare. Restrictions apply to the availability of these data, which were used under license for the current study, so supporting data are not publicly available.

Acknowledgments: The authors would like to thank Patrick Reilly for proofreading the article. We also wish to thank the County Council in Region Skåne for providing financial and administrative support to this study.

Conflicts of Interest: The authors declare no potential conflict of interest associated with this manuscript.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef]
2. Global Burden of Disease Cancer Collaboration; Fitzmaurice, C.; Abate, D.; Abbasi, N.; Abbastabar, H.; Abd-Allah, F.; Abdel-Rahman, O.; Abdelalim, A.; Abdoli, A.; Abdollahpour, I.; et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: A systematic analysis for the global burden of disease study. *JAMA Oncol.* **2019**, *5*, 1749–1768. [CrossRef]
3. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
4. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30. [CrossRef]
5. Lightowlers, R.N.; Chinnery, P.F.; Turnbull, D.M.; Howell, N. Mammalian mitochondrial genetics: Heredity, heteroplasmy and disease. *Trends Genet.* **1997**, *13*, 450–455. [CrossRef]
6. Veltri, K.L.; Espiritu, M.; Singh, G. Distinct genomic copy number in mitochondria of different mammalian organs. *J. Cell Physiol.* **1990**, *143*, 160–164. [CrossRef]
7. Yakes, F.M.; Van Houten, B. Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 514–519. [CrossRef] [PubMed]
8. Muftuoglu, M.; Mori, M.P.; de Souza-Pinto, N.C. Formation and repair of oxidative damage in the mitochondrial DNA. *Mitochondrion* **2014**, *17*, 164–181. [CrossRef]
9. Wu, S.; Li, X.; Meng, S.; Fung, T.; Chan, A.T.; Liang, G.; Giovannucci, E.; De Vivo, I.; Lee, J.H.; Nan, H. Fruit and vegetable consumption, cigarette smoke, and leukocyte mitochondrial DNA copy number. *Am. J. Clin. Nutr.* **2019**, *109*, 424–432. [CrossRef]
10. Fu, Y.; Tigano, M.; Sfeir, A. Safeguarding mitochondrial genomes in higher eukaryotes. *Nat. Struct. Mol. Biol.* **2020**, *27*, 687–695. [CrossRef]

11. Lee, H.C.; Wei, Y.H. Mitochondrial biogenesis and mitochondrial DNA maintenance of mammalian cells under oxidative stress. *Int. J. Biochem. Cell Biol.* **2005**, *37*, 822–834. [CrossRef]
12. Fontana, G.A.; Gahlon, H.L. Mechanisms of replication and repair in mitochondrial DNA deletion formation. *Nucleic Acids Res.* **2020**, *48*, 11244–11258. [CrossRef]
13. Lee, H.C.; Wei, Y.H. Mitochondrial role in life and death of the cell. *J. Biomed. Sci.* **2000**, *7*, 2–15. [CrossRef]
14. Shen, M.; Zhang, L.; Bonner, M.R.; Liu, C.S.; Li, G.; Vermeulen, R.; Dosemeci, M.; Yin, S.; Lan, Q. Association between mitochondrial DNA copy number, blood cell counts, and occupational benzene exposure. *Environ. Mol. Mutagen.* **2008**, *49*, 453–457. [CrossRef] [PubMed]
15. Anderson, A.P.; Luo, X.; Russell, W.; Yin, Y.W. Oxidative damage diminishes mitochondrial DNA polymerase replication fidelity. *Nucleic Acids Res.* **2020**, *48*, 817–829. [CrossRef] [PubMed]
16. Reznik, E.; Miller, M.L.; Senbabaoglu, Y.; Riaz, N.; Sarungbam, J.; Tickoo, S.K.; Al-Ahmadie, H.A.; Lee, W.; Seshan, V.E.; Hakimi, A.A.; et al. Mitochondrial DNA copy number variation across human cancers. *eLife* **2016**, *5*, e10769. [CrossRef] [PubMed]
17. Ashar, F.N.; Moes, A.; Moore, A.Z.; Grove, M.L.; Chaves, P.H.M.; Coresh, J.; Newman, A.B.; Matteini, A.M.; Bandeen-Roche, K.; Boerwinkle, E.; et al. Association of mitochondrial DNA levels with frailty and all-cause mortality. *J. Mol. Med.* **2015**, *93*, 177–186. [CrossRef] [PubMed]
18. Gaziev, A.I.; Abdullaev, S.; Podlutsky, A. Mitochondrial function and mitochondrial DNA maintenance with advancing age. *Biogerontology* **2014**, *15*, 417–438. [CrossRef]
19. Kim, J.H.; Im, J.A.; Lee, D.C. The relationship between leukocyte mitochondrial DNA contents and metabolic syndrome in postmenopausal women. *Menopause* **2012**, *19*, 582–587. [CrossRef]
20. Verhoeven, J.E.; Revesz, D.; Picard, M.; Epel, E.E.; Wolkowitz, O.M.; Matthews, K.A.; Penninx, B.; Puterman, E. Depression, telomeres and mitochondrial DNA: Between- and within-person associations from a 10-year longitudinal study. *Mol. Psychiatry* **2018**, *23*, 850–857. [CrossRef]
21. Czarny, P.; Wigner, P.; Strycharz, J.; Swiderska, E.; Synowiec, E.; Szatkowska, M.; Sliwinska, A.; Talarowska, M.; Szemraj, J.; Su, K.P.; et al. Mitochondrial DNA copy number, damage, repair and degradation in depressive disorder. *World J. Biol. Psychiatry* **2020**, *21*, 91–101. [CrossRef] [PubMed]
22. Chen, S.; Xie, X.; Wang, Y.; Gao, Y.; Xie, X.; Yang, J.; Ye, J. Association between leukocyte mitochondrial DNA content and risk of coronary heart disease: A case-control study. *Atherosclerosis* **2014**, *237*, 220–226. [CrossRef]
23. Zhang, Y.; Guallar, E.; Ashar, F.N.; Longchamps, R.J.; Castellani, C.A.; Lane, J.; Grove, M.L.; Coresh, J.; Sotoodehnia, N.; Ilkhanoff, L.; et al. Association between mitochondrial DNA copy number and sudden cardiac death: Findings from the Atherosclerosis risk in communities study (ARIC). *Eur. Heart J.* **2017**, *38*, 3443–3448. [CrossRef] [PubMed]
24. Fazzini, F.; Lamina, C.; Raftopoulou, A.; Koller, A.; Fuchsberger, C.; Pattaro, C.; Del Greco, F.M.; Dottelmayer, P.; Fendt, L.; Fritz, J.; et al. Association of mitochondrial DNA copy number with metabolic syndrome and type 2 diabetes in 14,176 individuals. *J. Intern. Med.* **2021**, *290*, 190–202. [CrossRef] [PubMed]
25. Asmann, Y.W.; Stump, C.S.; Short, K.R.; Coenen-Schimke, J.M.; Guo, Z.; Bigelow, M.L.; Nair, K.S. Skeletal muscle mitochondrial functions, mitochondrial DNA copy numbers, and gene transcript profiles in type 2 diabetic and nondiabetic subjects at equal levels of low or high insulin and euglycemia. *Diabetes* **2006**, *55*, 3309–3319. [CrossRef] [PubMed]
26. Sookoian, S.; Flichman, D.; Scian, R.; Rohr, C.; Dopazo, H.; Gianotti, T.F.; Martino, J.S.; Castano, G.O.; Pirola, C.J. Mitochondrial genome architecture in non-alcoholic fatty liver disease. *J. Pathol.* **2016**, *240*, 437–449. [CrossRef]
27. Malik, A.N.; Simoes, I.C.M.; Rosa, H.S.; Khan, S.; Karkucinska-Wieckowska, A.; Wieckowski, M.R. A diet induced maladaptive increase in hepatic mitochondrial DNA precedes OXPHOS defects and may contribute to non-alcoholic fatty liver disease. *Cells* **2019**, *8*, 1222. [CrossRef] [PubMed]
28. Fazzini, F.; Lamina, C.; Fendt, L.; Schultheiss, U.T.; Kotsis, F.; Hicks, A.A.; Meiselbach, H.; Weissensteiner, H.; Forer, L.; Krane, V.; et al. Mitochondrial DNA copy number is associated with mortality and infections in a large cohort of patients with chronic kidney disease. *Kidney Int.* **2019**, *96*, 480–488. [CrossRef]
29. Eirin, A.; Saad, A.; Tang, H.; Herrmann, S.M.; Woollard, J.R.; Lerman, A.; Textor, S.C.; Lerman, L.O. Urinary mitochondrial DNA copy number identifies chronic renal injury in hypertensive patients. *Hypertension* **2016**, *68*, 401–410. [CrossRef]
30. Yang, S.Y.; Castellani, C.A.; Longchamps, R.J.; Pillalamarri, V.K.; O'Rourke, B.; Guallar, E.; Arking, D.E. Blood-derived mitochondrial DNA copy number is associated with gene expression across multiple tissues and is predictive for incident neurodegenerative disease. *Genome Res.* **2021**, *31*, 349–358. [CrossRef]
31. Blackburn, E.H. Telomeres and telomerase: The means to the end (Nobel lecture). *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 7405–7421. [CrossRef] [PubMed]
32. Von Zglinicki, T. Oxidative stress shortens telomeres. *Trends Biochem. Sci.* **2002**, *27*, 339–344. [CrossRef]
33. Rudolph, K.L.; Millard, M.; Bosenberg, M.W.; DePinho, R.A. Telomere dysfunction and evolution of intestinal carcinoma in mice and humans. *Nat. Genet.* **2001**, *28*, 155–159. [CrossRef] [PubMed]
34. Kimura, M.; Gazitt, Y.; Cao, X.; Zhao, X.; Lansdorp, P.M.; Aviv, A. Synchrony of telomere length among hematopoietic cells. *Exp. Hematol.* **2010**, *38*, 854–859. [CrossRef]
35. Bonafe, M.; Sabbatinelli, J.; Olivieri, F. Exploiting the telomere machinery to put the brakes on inflamm-aging. *Ageing Res. Rev.* **2020**, *59*, 101027. [CrossRef] [PubMed]
36. Shay, J.W. Role of telomeres and telomerase in aging and cancer. *Cancer Discov.* **2016**, *6*, 584–593. [CrossRef]

37. Wu, X.; Amos, C.I.; Zhu, Y.; Zhao, H.; Grossman, B.H.; Shay, J.W.; Luo, S.; Hong, W.K.; Spitz, M.R. Telomere dysfunction: A potential cancer predisposition factor. *J. Natl. Cancer Inst.* **2003**, *95*, 1211–1218. [CrossRef]
38. Calado, R.T.; Young, N.S. Telomere diseases. *N. Engl. J. Med.* **2009**, *361*, 2353–2365. [CrossRef]
39. Satyanarayana, A.; Manns, M.P.; Rudolph, K.L. Telomeres and telomerase: A dual role in hepatocarcinogenesis. *Hepatology* **2004**, *40*, 276–283. [CrossRef]
40. Houben, J.M.; Moonen, H.J.; van Schooten, F.J.; Hageman, G.J. Telomere length assessment: Biomarker of chronic oxidative stress? *Free Radic. Biol. Med.* **2008**, *44*, 235–246. [CrossRef] [PubMed]
41. Liu, C.S.; Tsai, C.S.; Kuo, C.L.; Chen, H.W.; Lii, C.K.; Ma, Y.S.; Wei, Y.H. Oxidative stress-related alteration of the copy number of mitochondrial DNA in human leukocytes. *Free Radic. Res.* **2003**, *37*, 1307–1317. [CrossRef]
42. Hang, D.; Nan, H.; Kvaerner, A.S.; De Vivo, I.; Chan, A.T.; Hu, Z.; Shen, H.; Giovannucci, E.; Song, M. Longitudinal associations of lifetime adiposity with leukocyte telomere length and mitochondrial DNA copy number. *Eur. J. Epidemiol.* **2018**, *33*, 485–495. [CrossRef]
43. Samsioe, G.; Lidfeldt, J.; Nerbrand, C.; Nilsson, P. The women's health in the Lund area (WHILA) study—An overview. *Maturitas* **2010**, *65*, 37–45. [CrossRef]
44. Cawthon, R.M. Telomere measurement by quantitative PCR. *Nucleic Acids Res.* **2002**, *30*, e47. [CrossRef] [PubMed]
45. Wang, X.; Sundquist, K.; Hedelius, A.; Palmer, K.; Memon, A.A.; Sundquist, J. Leukocyte telomere length and depression, anxiety and stress and adjustment disorders in primary health care patients. *BMC Psychiatry* **2017**, *17*, 148. [CrossRef] [PubMed]
46. Memon, A.A.; Zoller, B.; Hedelius, A.; Wang, X.; Stenman, E.; Sundquist, J.; Sundquist, K. Quantification of mitochondrial DNA copy number in suspected cancer patients by a well optimized ddPCR method. *Biomol. Detect. Quantif.* **2017**, *13*, 32–39. [CrossRef] [PubMed]
47. Mi, J.; Tian, G.; Liu, S.; Li, X.; Ni, T.; Zhang, L.; Wang, B. The relationship between altered mitochondrial DNA copy number and cancer risk: A meta-analysis. *Sci. Rep.* **2015**, *5*, 10039. [CrossRef] [PubMed]
48. Zhu, X.; Han, W.; Xue, W.; Zou, Y.; Xie, C.; Du, J.; Jin, G. The association between telomere length and cancer risk in population studies. *Sci. Rep.* **2016**, *6*, 22243. [CrossRef]
49. Hofmann, J.N.; Hosgood, H.D., III; Liu, C.S.; Chow, W.H.; Shuch, B.; Cheng, W.L.; Lin, T.T.; Moore, L.E.; Lan, Q.; Rothman, N.; et al. A nested case-control study of leukocyte mitochondrial DNA copy number and renal cell carcinoma in the prostate, lung, colorectal and ovarian cancer screening trial. *Carcinogenesis* **2014**, *35*, 1028–1031. [CrossRef]
50. Kim, C.; Bassig, B.A.; Seow, W.J.; Hu, W.; Purdue, M.P.; Huang, W.Y.; Liu, C.S.; Cheng, W.L.; Mannisto, S.; Vermeulen, R.; et al. Mitochondrial DNA copy number and chronic lymphocytic leukemia/small lymphocytic lymphoma risk in two prospective studies. *Cancer Epidemiol. Biomark. Prev.* **2015**, *24*, 148–153. [CrossRef]
51. Lan, Q.; Lim, U.; Liu, C.S.; Weinstein, S.J.; Chanock, S.; Bonner, M.R.; Virtamo, J.; Albanes, D.; Rothman, N. A prospective study of mitochondrial DNA copy number and risk of non-Hodgkin lymphoma. *Blood* **2008**, *112*, 4247–4249. [CrossRef]
52. Hofmann, J.N.; Lan, Q.; Cawthon, R.; Hosgood, H.D., III; Shuch, B.; Moore, L.E.; Rothman, N.; Chow, W.H.; Purdue, M.P. A prospective study of leukocyte telomere length and risk of renal cell carcinoma. *Cancer Epidemiol. Biomark. Prev.* **2013**, *22*, 997–1000. [CrossRef]
53. Chen, N.; Wen, S.; Sun, X.; Fang, Q.; Huang, L.; Liu, S.; Li, W.; Qiu, M. Elevated mitochondrial DNA copy number in peripheral blood and tissue predict the opposite outcome of cancer: A meta-analysis. *Sci. Rep.* **2016**, *6*, 37404. [CrossRef] [PubMed]
54. Ennour-Idrissi, K.; Maunsell, E.; Diorio, C. Telomere length and breast cancer prognosis: A systematic review. *Cancer Epidemiol. Biomark. Prev.* **2017**, *26*, 3–10. [CrossRef]
55. Shanta, K.; Nakayama, K.; Ishikawa, M.; Ishibashi, T.; Yamashita, H.; Sato, S.; Sasamori, H.; Sawada, K.; Kurose, S.; Mahmud, H.M.; et al. Prognostic value of peripheral blood lymphocyte telomere length in gynecologic malignant tumors. *Cancers* **2020**, *12*, 1469. [CrossRef]
56. Kotsopoulos, J.; Prescott, J.; De Vivo, I.; Fan, I.; McLaughlin, J.; Rosen, B.; Risch, H.; Sun, P.; Narod, S.A. Telomere length and mortality following a diagnosis of ovarian cancer. *Cancer Epidemiol. Biomark. Prev.* **2014**, *23*, 2603–2606. [CrossRef]
57. Gandini, S.; Botteri, E.; Iodice, S.; Boniol, M.; Lowenfels, A.B.; Maisonneuve, P.; Boyle, P. Tobacco smoking and cancer: A meta-analysis. *Int. J. Cancer* **2008**, *122*, 155–164. [CrossRef] [PubMed]
58. Astuti, Y.; Wardhana, A.; Watkins, J.; Wulaningsih, W.; Network, P.R. Cigarette smoking and telomere length: A systematic review of 84 studies and meta-analysis. *Environ. Res.* **2017**, *158*, 480–489. [CrossRef]
59. Warburg, O. On the origin of cancer cells. *Science* **1956**, *123*, 309–314. [CrossRef] [PubMed]
60. Ding, J.; Sidore, C.; Butler, T.J.; Wing, M.K.; Qian, Y.; Meirelles, O.; Busonero, F.; Tsoi, L.C.; Maschio, A.; Angius, A.; et al. Assessing mitochondrial DNA variation and copy number in lymphocytes of ~2000 Sardinians using tailored sequencing analysis tools. *PLoS Genet.* **2015**, *11*, e1005306. [CrossRef]
61. Castellani, C.A.; Longchamps, R.J.; Sun, J.; Guallar, E.; Arking, D.E. Thinking outside the nucleus: Mitochondrial DNA copy number in health and disease. *Mitochondrion* **2020**, *53*, 214–223. [CrossRef] [PubMed]
62. Lee, H.C.; Wei, Y.H. Mitochondrial DNA instability and metabolic shift in human cancers. *Int. J. Mol. Sci.* **2009**, *10*, 674–701. [CrossRef]
63. Mengel-From, J.; Thinggaard, M.; Dalgard, C.; Kyvik, K.O.; Christensen, K.; Christiansen, L. Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. *Hum. Genet.* **2014**, *133*, 1149–1159. [CrossRef]

64. Guo, W.; Zheng, B.; Cai, Z.; Xu, L.; Guo, D.; Cao, L.; Wang, Y. The polymorphic AluYb8 insertion in the MUTYH gene is associated with reduced type 1 protein expression and reduced mitochondrial DNA content. *PLoS ONE* **2013**, *8*, e70718. [CrossRef] [PubMed]
65. Linkowska, K.; Jawien, A.; Marszalek, A.; Malyarchuk, B.A.; Tonska, K.; Bartnik, E.; Skonieczna, K.; Grzybowski, T. Mitochondrial DNA polymerase gamma mutations and their implications in mtDNA alterations in colorectal cancer. *Ann. Hum. Genet.* **2015**, *79*, 320–328. [CrossRef] [PubMed]
66. Cai, N.; Li, Y.; Chang, S.; Liang, J.; Lin, C.; Zhang, X.; Liang, L.; Hu, J.; Chan, W.; Kendler, K.S.; et al. Genetic control over mtDNA and its relationship to major depressive disorder. *Curr. Biol.* **2015**, *25*, 3170–3177. [CrossRef] [PubMed]
67. Wang, X.; Hart, J.E.; Liu, Q.; Wu, S.; Nan, H.; Laden, F. Association of particulate matter air pollution with leukocyte mitochondrial DNA copy number. *Environ. Int.* **2020**, *141*, 105761. [CrossRef] [PubMed]
68. Zhong, J.; Cayir, A.; Trevisi, L.; Sanchez-Guerra, M.; Lin, X.; Peng, C.; Bind, M.A.; Prada, D.; Laue, H.; Brennan, K.J.; et al. Traffic-related air pollution, blood pressure, and adaptive response of mitochondrial abundance. *Circulation* **2016**, *133*, 378–387. [CrossRef]
69. Hou, L.; Zhu, Z.Z.; Zhang, X.; Nordio, F.; Bonzini, M.; Schwartz, J.; Hoxha, M.; Dioni, L.; Marinelli, B.; Pegoraro, V.; et al. Airborne particulate matter and mitochondrial damage: A cross-sectional study. *Environ. Health* **2010**, *9*, 48. [CrossRef]
70. Young, M.J. Off-target effects of drugs that disrupt human mitochondrial DNA maintenance. *Front. Mol. Biosci.* **2017**, *4*, 74. [CrossRef]
71. Kopinski, P.K.; Singh, L.N.; Zhang, S.; Lott, M.T.; Wallace, D.C. Mitochondrial DNA variation and cancer. *Nat. Rev. Cancer* **2021**, *21*, 431–445. [CrossRef]
72. Jafri, M.A.; Ansari, S.A.; Alqahtani, M.H.; Shay, J.W. Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med.* **2016**, *8*, 69. [CrossRef]
73. Gaspar, T.B.; Sa, A.; Lopes, J.M.; Sobrinho-Simoes, M.; Soares, P.; Vinagre, J. Telomere maintenance mechanisms in cancer. *Genes* **2018**, *9*, 241. [CrossRef]
74. Shay, J.W.; Wright, W.E. Role of telomeres and telomerase in cancer. *Semin. Cancer Biol.* **2011**, *21*, 349–353. [CrossRef]
75. Willeit, P.; Willeit, J.; Mayr, A.; Weger, S.; Oberhollenzer, F.; Brandstatter, A.; Kronenberg, F.; Kiechl, S. Telomere length and risk of incident cancer and cancer mortality. *JAMA* **2010**, *304*, 69–75. [CrossRef]
76. Srinivas, N.; Rachakonda, S.; Kumar, R. Telomeres and telomere length: A general overview. *Cancers* **2020**, *12*, 558. [CrossRef]
77. Gong, Y.; Stock, A.J.; Liu, Y. The enigma of excessively long telomeres in cancer: Lessons learned from rare human POT1 variants. *Curr. Opin. Genet. Dev.* **2020**, *60*, 48–55. [CrossRef]
78. Zhu, Y.; Liu, X.; Ding, X.; Wang, F.; Geng, X. Telomere and its role in the aging pathways: Telomere shortening, cell senescence and mitochondria dysfunction. *Biogerontology* **2019**, *20*, 1–16. [CrossRef]
79. Hu, J.; Hwang, S.S.; Liesa, M.; Gan, B.; Sahin, E.; Jaskelioff, M.; Ding, Z.; Ying, H.; Boutin, A.T.; Zhang, H.; et al. Antitelomerase therapy provokes ALT and mitochondrial adaptive mechanisms in cancer. *Cell* **2012**, *148*, 651–663. [CrossRef] [PubMed]
80. Gallicchio, L.; Gadalla, S.M.; Murphy, J.D.; Simonds, N.I. The effect of cancer treatments on telomere length: A systematic review of the literature. *J. Natl. Cancer Inst.* **2018**, *110*, 1048–1058. [CrossRef] [PubMed]
81. Bao, D.; Ba, Y.; Zhou, F.; Zhao, J.; Yang, Q.; Ge, N.; Guo, X.; Wu, Z.; Zhang, H.; Yang, H.; et al. Alterations of telomere length and mtDNA copy number are associated with overall survival in hepatocellular carcinoma patients treated with transarterial chemoembolization. *Cancer Chemother. Pharmacol.* **2016**, *78*, 791–799. [CrossRef] [PubMed]
82. Antoun, S.; Atallah, D.; Tahtouh, R.; Assaf, M.D.; Moubarak, M.; Ayoub, E.N.; Chahine, G.; Hilal, G. Glucose restriction combined with chemotherapy decreases telomere length and cancer antigen-125 secretion in ovarian carcinoma. *Oncol. Lett.* **2019**, *19*, 1338–1350. [CrossRef]

Paper II

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 Ashfaque A. Memon[a]

[a]Center for Primary Health Care Research, Lund University/Region Skåne, Malmö, Sweden
[b]Department of Family Medicine and Community Health, Icahn School of Medicine at Mount Sinai, New York, United States
[c]Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, United States
[d]Center for Community-Based Healthcare Research and Education (CoHRE), Department of Functional Pathology, School of Medicine, Shimane University, Matsue, Japan

eBioMedicine
2023;88: 104432
Published Online 10 January 2023
https://doi.org/10.1016/j.ebiom.2022.104432

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-eqtls.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).

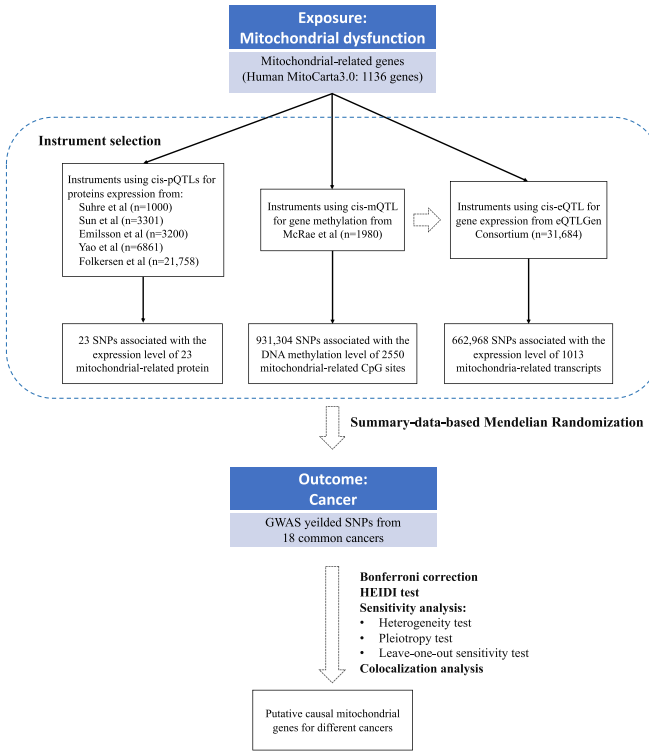


Fig. 1: **Flowchart of the analyses performed.**

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 ($P_{SMR}<4.936 \times 10^{-5}$ [Bonferroni correction, $P < 0.05/1013$]) followed by the HEIDI test ($P_{HEIDI}>0.01$) 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

Type of cancer	Gene	Instrumental variant	eQTL size effect[a]	β coefficient	OR (95% CI)[b]	P_{SMR} value	Colocalization[c]
Breast cancer							
	FDPS	rs6677385	-0·09	-0·42	0·66 (0·49–0·83)	$9·77\times10^{-7}$	0·89
	HSCB	rs6519752	0·08	0·58	1·78 (1·52–2·05)	$1·57\times10^{-5}$	0·74
	MPC2	rs203777	-0·03	0·33	1·39 (1·25–1·54)	$1·08\times10^{-5}$	0·004
	MRPS18C	rs1565909	0·14	0·29	1·34 (1·25–1·43)	$1·09\times10^{-10}$	<0·001
	NSUN4	rs41293273	0·03	0·05	1·05 (1·03–1·07)	$5·24\times10^{-6}$	0·90
	PARK7	rs35675666	-0·04	-0·09	0·91 (0·87–0·95)	$2·15\times10^{-6}$	0·003
	YBEY	rs62224180	0·78	0·04	1·04 (1·02–1·06)	$3·05\times10^{-5}$	<0·001
Prostate cancer							
	CASP8	rs7560328	-0·21	0·20	1·22 (1·15–1·30)	$2·89\times10^{-7}$	<0·001
	NSUN4	rs41293273	0·66	0·06	1·06 (1·03–1·09)	$1·01\times10^{-5}$	0·94
	SLC25A37	rs2928682	-0·49	0·13	1·13 (1·10–1·17)	$2·84\times10^{-12}$	<0·001
	UQCC1	rs4911178	-0·46	0·09	1·09 (1·06–1·13)	$8·70\times10^{-7}$	0·004
Gastric cancer							
	BAK1	rs210143	0·67	-0·14	0·87 (0·81–0·92)	$6·55\times10^{-7}$	<0·001

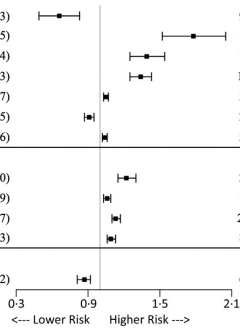


Fig. 2: **Mendelian randomization results for the association between the expression of mitochondrial genes and cancer risk.** [a]Represents the effect size (β) of a variant on mRNA expressions. β > 0 means positive association, and β < 0 means negative association. [b]Odd ratios were calculated by the expectation of causal estimate (β coefficient). [c]'Colocalization' 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

Type of cancer	Gene	Methylation probe	Instrumental variant	mQTL size effect[a]	β coefficient	OR (95% CI)[b]	P_{SMR} value	Colocalization[c]
Breast cancer								
	BCL2L11	cg09907170	rs73954926	0·82	0·07	1·07 (1·04–1·11)	$1·08\times10^{-5}$	0·003
		cg27608154	rs73954941	-1·51	-0·04	0·96 (0·95–0·98)	$2·56\times10^{-6}$	0·003
	MRPL23	cg07977153	rs4929956	0·26	-0·16	0·85 (0·78–0·92)	$1·32\times10^{-7}$	<0·001
	NSUN4	cg14993813	rs6682266	-0·35	-0·09	0·91 (0·87–0·96)	$1·50\times10^{-5}$	0·96
		cg17875957	rs5013329	-0·40	-0·08	0·92 (0·89–0·96)	$9·86\times10^{-6}$	0·96
		cg17806798	rs41293277	-1·12	-0·03	0·97 (0·96–0·99)	$5·34\times10^{-6}$	0·96
		cg04241075	rs41293277	-1·04	-0·03	0·97 (0·96–0·98)	$5·59\times10^{-6}$	0·96
		cg06741803	rs56063031	-0·36	-0·09	0·92 (0·88–0·96)	$1·56\times10^{-5}$	0·96
		cg00530320	rs6681857	0·47	0·07	1·07 (1·04–1·10)	$5·81\times10^{-6}$	0·96
		cg15580309	rs111226885	1·19	0·03	1·03 (1·02–1·04)	$1·72\times10^{-6}$	0·96
	SLC25A22	cg23587532	rs7928917	0·23	-0·16	0·85 (0·78–0·92)	$6·55\times10^{-6}$	<0·001
		cg11475788	rs61876744	0·21	-0·19	0·83 (0·75–0·91)	$8·67\times10^{-6}$	<0·001
	SLC25A44	cg19263494	rs72708291	-0·84	-0·04	0·96 (0·95–0·98)	$6·58\times10^{-6}$	<0·001
	TRMT1	cg22237401	rs74569397	0·66	0·12	1·12 (1·08–1·17)	$1·40\times10^{-6}$	0·007
		cg12014333	rs16995252	0·59	0·13	1·14 (1·08–1·19)	$3·22\times10^{-6}$	0·007
Prostate cancer								
	BIK	cg07972488	rs2294413	0·20	0·40	1·48 (1·31–1·65)	$2·76\times10^{-6}$	<0·001
	NSUN4	cg04241075	rs41293277	-1·04	-0·04	0·96 (0·95–0·98)	$1·16\times10^{-5}$	0·96
		cg17806798	rs41293277	-1·12	-0·03	0·97 (0·95–0·98)	$1·11\times10^{-5}$	0·96
		cg15580309	rs111226885	1·19	0·03	1·03 (1·02–1·05)	$7·95\times10^{-6}$	0·96
	NUDT5	cg22687873	rs4750175	0·03	-0·04	0·96 (0·95–0·98)	$1·15\times10^{-5}$	0·99
	TSPO	cg13160331	rs138909	0·19	0·34	1·40 (1·26–1·54)	$2·72\times10^{-6}$	<0·001
	VARS2	cg02186769	rs2233959	0·24	-0·18	0·84 (0·76–0·92)	$1·63\times10^{-5}$	0·009
		cg05103231	rs2524108	0·24	-0·22	0·80 (0·71–0·90)	$1·24\times10^{-5}$	<0·001
		cg12457901	rs2233980	0·86	-0·07	0·93 (0·90–0·96)	$2·19\times10^{-6}$	0·009
		cg16113650	rs2249935	0·20	-0·36	0·70 (0·54–0·86)	$1·24\times10^{-5}$	<0·001
		cg26467571	rs9264490	0·28	-0·19	0·83 (0·77–0·91)	$2·34\times10^{-6}$	<0·001
Gastric cancer								
	BAK1	cg00700324	rs511515	-0·28	0·34	1·40 (1·25–1·55)	$1·41\times10^{-5}$	<0·001
Lung cancer								
	VARS2	cg12457901	rs2523593	0·84	0·20	1·22 (1·14–1·30)	$1·18\times10^{-6}$	0·98
		cg14935711	rs2523593	0·76	0·22	1·24 (1·15–1·33)	$1·41\times10^{-6}$	0·98
		cg15848685	rs2596500	0·80	0.20	1·23 (1·14–1·31)	$2·81\times10^{-6}$	0·98
		cg16958594	rs2596495	0·75	0·20	1·22 (1·13–1·31)	$7·23\times10^{-6}$	0·98
Melanoma								
	SPG7	cg09560549	rs8060502	0·39	0·003	1·003 (1·002–1·004)	$2·21\times10^{-8}$	<0·001
		cg15206445	rs8060502	0·44	0·003	1·003 (1·002–1·004)	$6·88\times10^{-9}$	<0·001

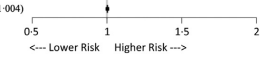


Fig. 3: **Mendelian randomization results for the association between mitochondrial gene methylations and cancer risk.** [a]Represents the effect size (β) of a variant on DNA methylation. β > 0 means positive association, and β < 0 means negative association. [b]Odd ratios were calculated by the expectation of causal estimate (β coefficient). [c]'Colocalization' 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 GWASATLAS[19] and PhenoScanner[20] 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 *VARS2* 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) *VARS2* 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 resistance[26] 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 m^5C 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 *VARS2* 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 *VARS2* as a causal gene only for lung cancer. The mechanisms of *VARS2* 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

1 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.
2 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.
3 Miwa S, Kashyap S, Chini E, von Zglinicki T. Mitochondrial dysfunction in cell senescence and aging. *J Clin Invest*. 2022;132(13):e158447.
4 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.
5 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.
6 Moro L. Mitochondrial DNA and MitomiR variations in pancreatic cancer: potential diagnostic and prognostic biomarkers. *Int J Mol Sci*. 2021;22(18).
7 Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. 2018;362:k601.
8 Lawlor DA. Commentary: two-sample Mendelian randomization: opportunities and challenges. *Int J Epidemiol*. 2016;45(3):908–915.
9 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.
10 McRae AF, Marioni RE, Shah S, et al. Identification of 55,000 replicated DNA methylation QTL. *Sci Rep*. 2018;8(1):17605.
11 Cuschieri S. The STROBE guidelines. *Saudi J Anaesth*. 2019;13(Suppl 1):S31–S34.
12 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.
13 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.
14 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.
15 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.
16 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.
17 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.
18 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.
19 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.
20 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.
21 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.
22 Sun S, McKenna CE. Farnesyl pyrophosphate synthase modulators: a patent review (2006 - 2010). *Expert Opin Ther Pat*. 2011;21(9):1433–1451.
23 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.
24 Seshacharyulu P, Rachagani S, Muniyan S, et al. FDPS cooperates with PTEN loss to promote prostate cancer progression through modulation of small GTPases/AKT axis. *Oncogene*. 2019;38(26):5265–5280.
25 Clendening JW, Pandyra A, Boutros PC, et al. Dysregulation of the mevalonate pathway promotes transformation. *Proc Natl Acad Sci U S A*. 2010;107(34):15051–15056.
26 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.
27 Souchek 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.
28 Berndt N, Hamilton AD, Sebti SM. Targeting protein prenylation for cancer therapy. *Nat Rev Cancer*. 2011;11(11):775–791.
29 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.
30 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.
31 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.
32 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.
33 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.
34 Tong XY, Liao X, Gao M, et al. Identification of NUDT5 inhibitors from approved drugs. *Front Mol Biosci*. 2020;7:44.
35 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.
36 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.
37 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.
38 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.

Printed by Media-Tryck, Lund 2023 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences, Malmö

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2023:113
ISBN 978-91-8021-454-4
ISSN 1652-8220