Pre-diagnostic Leukocyte Mitochondrial DNA Copy Number and Colorectal Cancer Risk

Keming Yang[1], Xin Li[1], Michele R. Forman[2], Patrick O. Monahan[3], Bret H. Graham[4], Amit Joshi[5], Mingyang Song[5,6,7], Dong Hang[7,8], Shuji Ogino[6,9,10], Edward L. Giovannucci[6,7,11], Immaculata De Vivo[6,11], Andrew T. Chan[5,11*], Hongmei Nan[1,12*]

Author affiliations:
1 - Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN, USA;
2 - Department of Nutrition Science, College of Health and Human Science, Purdue Center for Cancer Research, Purdue University, West Lafayette, IN, USA;
3 - Department of Biostatistics, School of Medicine and Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN, USA;
4 - Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA;
5 - Clinical and Translational Epidemiology Unit and Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA;
6 - Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA;
7 - Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA;
8 - Department of Epidemiology and Biostatistics, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, School of Public Health, Nanjing Medical University, Nanjing, China;
9 - Program in MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA, USA;
10 - Broad Institute of MIT and Harvard, Cambridge, MA, USA;
11 - Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA;
12 - IU Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, Indiana, USA

*** Co-senior/corresponding authors:**

Dr. Hongmei Nan: Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, R3-C216A, 980 W Walnut street, Indianapolis, IN 46202, USA
Telephone: 317-278-3907 E-mail: hnan@iu.edu

Dr. Andrew T. Chan: Clinical and Translational Epidemiology Unit and Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, GRJ-722, 55 Fruit St, Boston, MA 02114, USA
Telephone: 617-726-2862 E-mail: ACHAN@mgh.harvard.edu

This is the author's manuscript of the article published in final edited form as:

Yang, K., Li, X., Forman, M. R., Monahan, P. O., Graham, B. H., Joshi, A., … Nan, H. (n.d.). Pre-diagnostic Leukocyte Mitochondrial DNA Copy Number and Colorectal Cancer Risk. Carcinogenesis. https://doi.org/10.1093/carcin/bgz159

ABSTRACT

Mitochondrial DNA (mtDNA) is susceptible to oxidative stress and mutation. Few epidemiological studies have assessed the relationship between mtDNA copy number (mtDNAcn) and risk of colorectal cancer (CRC), with inconsistent findings. In this study, we examined the association between pre-diagnostic leukocyte mtDNAcn and CRC risk in a case-control study of 324 female cases and 658 matched controls nested within the Nurses' Health Study (NHS). Relative mtDNAcn in peripheral blood leukocytes was measured by quantitative PCR-based assay. Conditional logistic regression models were applied to estimate odds ratios (ORs) and 95% confidence intervals (95% CIs) for the association of interest. Results showed lower log-mtDNAcn was significantly associated with increased risk of CRC, in a dose-dependent relationship (*P* for trend < 0.0001). Compared to the 4th quartile, multivariable-adjusted OR (95% CI) was 1.10 (0.69, 1.76) for the 3rd quartile, 1.40 (0.89, 2.19) for the 2nd quartile, and 2.19 (1.43, 3.35) for the 1st quartile. In analysis by anatomic subsite of CRC, we found a significant inverse association for proximal colon cancer [lowest vs. highest quartile, multivariable-adjusted OR (95% CI) = 3.31 (1.70, 6.45), *P* for trend = 0.0003]. Additionally, stratified analysis according to the follow-up time since blood collection showed that the inverse association between mtDNAcn and CRC remained significant among individuals with ≥ 5 years' follow-up, and marginally significant among those with ≥ 10 years' follow-up since mtDNAcn testing, suggesting that mtDNAcn may serve as a long-term predictor for risk of CRC. In conclusion, pre-diagnostic leukocyte mtDNAcn was inversely associated with CRC risk. Further basic experimental studies are needed to explore the underlying biological mechanisms linking mtDNAcn to CRC carcinogenesis.

Keywords: mitochondrial DNA copy number; colorectal cancer;

Summary: Pre-diagnostic leukocyte mitochondrial DNA copy number, a reflection of oxidative stress damage, was inversely associated with risk of colorectal cancer, and may be a long-term predictor of colorectal cancer risk.

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer among men and women in the U.S.(1). Across worldwide, CRC is the third most common cancer in men and the second most common cancer in women (2). The incidence and mortality of CRC have been decreasing over the past decade, likely due in part to the successful implementation of screening programs (3). However, the disease is still the third leading cause of cancer death in the U.S. and the fourth across the world among men and women (1,2), posing an enormous health and socioeconomic burden (4). Thus, identifying biomarkers for CRC risk that might inform prevention and early diagnosis is of great public health importance.

Mitochondria are essential eukaryotic organelles containing their own genome, i.e., mitochondrial DNA (mtDNA), which is usually maternally inherited (5). MtDNA consists of approximately 16,569 bp double-stranded circular DNA and encodes only thirty-seven genes. Most mammalian cells contain between hundreds and over a thousand mitochondria per cell, and each mitochondrion has two to ten copies of mtDNA (5). Compared to nuclear DNA, mtDNA has a higher mutation rate and is particularly susceptible to oxidative stress, probably due to its proximity to the source of reactive oxygen species (ROS) and its lack of protective histones (6). Though the study of mtDNA repair pathways has lagged behind inquiries into nuclear DNA repair mechanisms, research has not only shown the existence of robust damage tolerance mechanisms in mitochondria, but also proposed various mtDNA repair pathways that may properly maintain the mitochondrial genome (7).

CRC is a heterogeneous disease associated with environmental and genetic factors through complicated interactions (8,9). Oxidative stress triggered by ROS may initiate and promote carcinogenesis (including colorectal carcinogenesis) by inducing inflammation, DNA

damage, gene mutations, and genomic instability (10,11). Because mtDNA copy number (mtDNAcn) is a major biomarker for oxidative DNA damage and mitochondrial dysfunction, it has been hypothesized that altered pre-diagnostic leukocyte mtDNAcn may be associated with risk of developing cancers, including CRC.

The few epidemiological studies that have assessed the relationship between mtDNAcn and risk of CRC have yielded inconsistent findings (12-14). A retrospective case-control study conducted by Qu *et al.* in a hospital setting in China first reported a positive association between mtDNAcn and CRC risk (12). Later, prospective case-control studies nested within the Shanghai Women's Health Study (SWHS) (13) reported an inverse association, and the Singapore Chinese Health Study (SCHS) (14) reported a U-shaped relationship. Evidence supporting the relationship between mtDNAcn and CRC risk in western populations has been lacking. Therefore, in this study we examined the association between pre-diagnostic leukocyte mtDNAcn and the risk of CRC in a case-control study of 324 CRC cases and 658 matched healthy controls nested within the Nurses' Health Study (NHS), a long-term prospective cohort study of women in the US.

METHODS

Study population

The Nurses' Health Study was initiated in 1976, when 121,700 female US registered nurses aged 30-55 years completed and returned questionnaires regarding their medical histories and baseline lifestyles. Biennially, participants completed self-administered follow-up questionnaires with updated information on their dietary habits and other lifestyle factors, medical history, and disease diagnosis. In 1989-1990, a total of 32,826 participants in the NHS provided blood samples. Details of the NHS have been previously published (15).

Colorectal cancer case ascertainment and control selection

CRC diagnoses were based on the self-report by nurses on biennial questionnaires and then confirmed by a pathologist. All CRC cases were incident cases diagnosed after blood collection. In this nested case-control study, we randomly selected 1-3 controls from the same cohort (NHS) of participants who were free of cancer (excluding non-melanoma skin cancer) up to and including the questionnaire cycle in which the case was diagnosed. Control subjects were matched to each case based on year of birth (±1 year), race, and fasting status at blood collection. A total of 324 CRC cases and 658 healthy controls were included. The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required.

Assessment of mtDNAcn

Details of the ascertainment and validation of leukocyte mtDNAcn for blood samples from the NHS participants and quality control procedures were described previously (16-18). More detail can be found in **Supplementary Methods**. Briefly, genomic DNA was extracted from the buffy-coat leukocytes in peripheral blood, according to the QIAmp (Qiagen, CA, USA) 96-spin blood protocol. Concentrations of DNA were measured by pico-green quantitation utilizing a Molecular Devices 96-well spectrophotometer. The quantitative Polymerase Chain Reaction (qPCR)-based assay was used to determine the ratio of the copy numbers of mitochondrial *ND2* gene to genomic single-copy gene (*AluYb8)* (N/S), which is proportional to the average number of mtDNAcn. The relative N/S ratio was then calculated by subtracting the N/S ratio of the calibrator DNA from the N/S ratio of each sample. The value of mtDNAcn was calculated as the exponentiated N/S ratio. Each sample was assayed in triplicate, and 10% replicate quality-control

(QC) samples were included. The coefficients of variation (CVs) for *ND2* and *AluYb8* were less than 1% among QC samples.

Assessment of covariates

Covariate data were collected through self-administered questionnaires at baseline (1976) and during follow-up biennially. In this study, we used covariates data from the questionnaire cycle closest to blood collection (1989-1990), including body mass index (BMI) calculated as height (m)/weight (kg)^2, smoking status, alcohol consumption, Alternate Healthy Eating Index (AHEI), physical activity, family history of CRC, regular use of aspirin and other non-steroidal anti-inflammatory drugs (NSAIDs), parity as well as menopausal status and postmenopausal hormone use. Specifically, height and weight were collected at baseline and then weight was updated at each biennial questionnaire; we used height at baseline and weight at blood collection to calculate BMI. We also included weight change from blood collection until two years before diagnosis of the cases and same cycle of their matched controls into the multivariable model; to minimize reverse causality of CRC on weight, we excluded the two years before diagnosis for the calculation of weight change. Participants who used aspirin (either standard or low-dose) at least 2 times/week on average were classified as regular aspirin users. Regular users of non-aspirin nonsteroidal anti-inflammatory drugs (NSAIDs) were participants who responded "yes" to regular use on questionnaires, defined as at least 2 times/week (19). Physical activity was represented by metabolic equivalent (MET)-hours/week. Specifically, each activity was assigned a MET value, which refers to the metabolic rates for each specific activity divided by metabolic rates at rest. MET-hours/week for each activity was calculated by multiplying average time per week in each activity by the MET of each activity. Then total MET-hours per week was derived by summing up MET-hours per week for each activity (20). Alternate Healthy Eating Index

(AHEI) is a dietary score (0-100 points) measuring the adherence to a dietary pattern characterized by foods and nutrients most predictive of risk of diseases; higher AHEI indicates healthier dietary quality. Foods/nutrients involved in AHEI development include whole grains, sugar-sweetened beverages, fruit juice, vegetables, fruits, nuts and legumes, red and processed meat, poly-unsaturated fatty acids, trans fats, long-chain (n-3) fats (EPA + DHA), and sodium (21). AHEI used in the current study was derived from the food frequency questionnaire at blood collection. The validity and reproducibility of physical activity and dietary information from the food frequency questionnaire have been reported elsewhere (22,23).

Statistical analysis

Log-transformed mtDNA copy numbers [log (mtDNAcn)] of cases and controls were stratified into four categories based on the quartiles of log (mtDNAcn) among all controls. Conditional logistic regression was applied to estimate the odds ratio (OR) and 95% confidence interval (CI) for the association of log (mtDNAcn) with CRC risk. Two models were analyzed: Model 1, the crude model without covariate adjustment; and Model 2, the multivariable model that adjusted for potential confounders, including BMI (in tertiles: 0-23.2 kg/m2, 23.2-26.6 kg/m2, ≥ 26.6 kg/m^2), physical activity (in tertiles, 0-8.2, 8.2-20.2, ≥ 20.2 MET-hours/week), smoking status (never, former, or current smokers), alcohol consumption (in tertiles, 0-0.8 g, 0.8-5.8 g, ≥ 5.8 g per day), menopausal status and postmenopausal hormone use (premenopausal, never and former users, current users), parity (0/1/2/3+children), Alternate healthy eating index (AHEI) (in tertiles, 0-42.8, 42.8-51.3, ≥ 51.3), regular aspirin use (yes/no), regular non-aspirin NSAIDs use (yes/no), and family history of colorectal cancer (yes/no) at blood collection, and weight change from blood collection until 2 years before diagnosis (in tertiles, < 0 kg, 0-2.72 kg, ≥ 2.72 kg). Conditional logistic regression models were also used to estimate the association

between mtDNAcn and CRC risk by anatomic subsites, including colon cancer (proximal, distal) and rectal cancer. Unconditional logistic regression with adjustment for matching factors and covariates was employed to further examine the effect of two-way interactions between mtDNAcn and potential confounders on the risk of CRC. The statistical significance of interaction was assessed using likelihood ratio test for cross-product terms of covariates and log (mtDNAcn).

To minimize the reverse influence of potential undetectable tumors if any at blood collection on mtDNAcn, we conducted a sensitivity analysis by removing cases diagnosed within 1 and 2 years after blood collection and their matched controls. In addition, to exclude any potential influence of colorectal polyps and inflammatory bowel disease (IBD) on mtDNAcn (especially among controls), we performed another sensitivity analysis by removing cases and controls who had colorectal polyps and/or IBDs before the time of CRC diagnosis. We also examined the association between mtDNAcn and CRC risk stratified by the follow-up time since blood collection to explore whether mtDNAcn has potential as a long-term predictive biomarker for risk of CRC. All statistical analyses were performed using SAS software, version 9.4 for UNIX (SAS Institute, North Carolina). All tests were two-sided and $P<0.05$ was considered statistically significant.

RESULTS

Basic characteristics of CRC cases (n = 324) and matched controls (n= 658) in this nested case-control study are presented in **Table 1**. Briefly, the mean age (standard deviation, SD) at blood collection for cases was 58.9 (6.7) years and for controls was 59.3 (6.6) years. Mean age (SD) at CRC diagnosis was 67.4 (7.5) years among cases. MtDNAcn was lower in cases than

controls. Compared to controls, relatively fewer cases were regular users of aspirin or NSAIDs, or current users of postmenopausal hormones, while relatively more cases were current smokers, had family history of CRC, and consumed higher amounts of alcohol. We also present those basic characteristics according to mtDNAcn quartiles after age standardization among 658 control subjects **(Table 2)**. Briefly, compared to the women in the highest quartile, percentages of current smokers and participants with family history of CRC were higher, while levels of total physical activity (MET-hours/week) were lower among the women in the lowest quartile of mtDNAcn.

For the association between mtDNAcn and CRC risk, we found that lower log-mtDNAcn level was significantly associated with an increased risk of CRC, with a dose-dependent relationship in both the crude and multivariable-adjusted models; compared to the crude model, results did not change materially after adjusting for a list of covariates **(Table 3).** Compared to the highest (4th) quartile, multivariable-adjusted OR (AOR, 95% CI) was 1.10 (0.69, 1.76) for the 3rd quartile, 1.40 (0.89, 2.19) for the 2nd quartile, and 2.19 (1.43, 3.35) for the 1st quartile (*P* for trend < 0.0001). In the further analysis of CRC by anatomic subsite, we observed a significant inverse association for proximal colon cancer [lowest vs. highest quartile, AOR (95% CI) =3.31 (1.70, 6.45), *P* for trend = 0.0003] **(Table 3)**. The inverse association was not statistically significant for distal colon cancer and rectal cancer, which may be due to the small number of cases with cancer at those subsites.

In the further sensitivity analysis to test any potential influence of colorectal polyps and inflammatory bowel disease (IBD) on mtDNAcn/CRC, the inverse association remained significant after removing cases and controls who had colorectal polyps and/or IBD before CRC diagnosis [lowest vs. highest quartile, AOR (95% CI) = 2.51 (1.52, 4.13), *P* for trend < 0.0001)].

In another sensitivity analysis examining the possible reverse influence of potential undetectable tumors (if any) at blood collection on mtDNAcn, the results did not change materially after removal of cases diagnosed within 1 and 2 years after blood collection and their matched controls, indicating minimal reverse causation [follow-up ≥ 1 year, lowest vs. highest quartile, AOR (95% CI) = 2.08 (1.33, 3.24), *P* for trend = 0.0006; follow-up ≥ 2 year, lowest vs. highest quartile, AOR (95% CI) = 1.92 (1.22, 3.02), *P* for trend = 0.004].

Moreover, we performed a stratified analysis according to the follow-up time since blood collection (i.e., mtDNAcn testing) **(Table 4)**. The inverse association between mtDNAcn and CRC remained significant among individuals with ≥ 5 years' follow-up since mtDNAcn testing [lowest vs. highest quartile, AOR (95% CI) = 1.98 (1.18, 3.33), *P* for trend = 0.009]. The inverse association was also marginally significant among those with ≥ 10 years' follow-up [lowest vs. highest quartile, AOR (95% CI) = 1.92 (0.94, 3.95), *P* for trend = 0.06]. These data suggest that mtDNAcn could serve as a long-term predictive marker for the risk of CRC.

We also examined the effect of interactions between mtDNAcn and potential confounders on the risk of CRC. We observed an effect modification of Alternate Healthy Eating Index (AHEI) on the association between mtDNAcn and CRC risk (*P* for interaction = 0.03). In the stratified analysis by AHEI, a significant inverse association between mtDNAcn and CRC risk was shown among individuals in the lowest AHEI (i.e., less healthy diet) tertile group [lowest vs. highest mtDNAcn quartile, AOR (95% CI) = 3.79 (1.77, 8.13), *P* for trend = 0.001]; the inverse associations were weaker and not statistically significant among those in the 2nd AHEI tertile group [lowest vs. highest mtDNAcn quartile, AOR (95% CI) = 1.62 (0.79, 3.35), *P* for trend = 0.11] and 3rd AHEI tertile group [AOR (95% CI) =1.75 (0.82, 3.74), *P* for trend = 0.10]. No significant effect modification appeared for other potential confounders, including

BMI, physical activity, weight change from blood collection until two years before diagnosis, smoking status, alcohol consumption, postmenopausal hormone use, parity, regular aspirin and non-aspirin NSAID use, and CRC family history (*P* for interactions >0.05) (data not shown).

DISCUSSION

In our nested case-control study, we report that pre-diagnostic leukocyte mtDNAcn was inversely associated with subsequent CRC risk in a dose-dependent manner. Our findings are in line with results from a case-control study of 444 CRC cases (mean baseline age = 58.6) and 1,423 controls (mean baseline age = 55.2) nested within the Shanghai Women's Health Study, in which Huang *et al.* found that lower mtDNAcn was associated with higher risk of CRC [lowest vs. highest tertile, OR (95% CI)=1.44 (1.06-1.94), *P for trend* = 0.02] (13). In another case-control study nested within the Singapore Chinese Health Study of women and men, Thyagarajan *et al.* reported a U-shaped relationship between mtDNAcn and CRC risk among 422 cases (mean baseline age = 66.1) and 874 controls (mean baseline age = 57.6) [lowest vs. 2nd quartile, OR = 1.81 (1.13-2.89), highest vs. 2nd quartile, OR = 3.40 (2.15-5.36), *P for curvilinearity* < 0.0001] (14).

Besides CRC, several other cancers have also been inversely associated with mtDNAcn in epidemiological studies. For example, Meng *et al.* studied the association between mtDNAcn and melanoma in a case-control study (272 cases and 293 controls) nested within the NHS, and found an inverse association among the high cumulative UV exposure group [low vs. high mtDNAcn, OR (95% CI) = 3.40 (1.46-7.92), *P* for trend=0.004] (16). In another study by Meng *et al.* using both NHS (women) and the Health Professionals Follow-Up Study (HPFS, men), among current smokers, those with median mtDNAcn levels were found to have higher risk of

lung cancer than those with high mtDNAcn levels [median vs. high mtDNAcn, OR (95% CI) = 2.09 (1.12-3.90)] (17). Also, Xie *et al.* found an inverse association between mtDNAcn and soft tissue sarcoma among 325 patients and 330 healthy controls (age, sex, ethnicity matched); among both men and women, lower mtDNAcn was associated with a significantly increased risk of soft tissue sarcoma [< median vs. ≥ median, AOR (95% CI) = 2.71(1.94–3.82)] (24). However, mixed results (including both positive and null associations) were also reported for the relationship between mtDNAcn and other cancers, such as renal cell carcinoma and non-Hodgkin lymphoma (25,26). Considering the complexity of carcinogenesis, it is possible that the relationship between mtDNAcn and cancer risk may be site-specific, depending on the specific organ or tissue of origin.

Elevated oxidative stress may affect the abundance of mitochondria and mtDNAcn as well as mitochondrial function (27,28). Recent evidence has shown the existence of various DNA-repair pathways in mitochondria, such as mismatch repair, base excision repair, homologous recombination and non-homologous end joining, lesion bypass, and mtDNA degradation (6,7). However, when the rate of oxidative damage overwhelms the ability of these mechanisms to repair mtDNA efficiently, mtDNA may proliferate, followed by the eventual loss of mtDNA (27). Specifically, when mtDNA is impaired by excessive oxidative stress, healthy mitochondria may first increase their DNA copy number to counteract the metabolic defects in injured mitochondria (27). However, when the damage exceeds the limitation of the feedback mechanism, increasing mtDNAcn can no longer cope with the stress. This results in a net decrease in mtDNAcn, because mtDNA undergoes degradation by the inner cellular enzyme system to prevent excessive accumulation of oxidative stress damage (28). These mechanisms may explain the inconsistent findings from previous studies, and the inverse association we

observed might be because extensive oxidative stress may have surpassed mitochondrial capacity to compensate for oxidative damage. Moreover, previous studies have demonstrated a strong positive association between mtDNAcn and telomere length (18,29), a crucial marker of cellular aging and the cumulative burden of oxidative stress (30). With each cell division telomeres undergo shortening, and oxidative stress could increase this erosion (31). The positive correlation between mtDNAcn and telomere length also implies the potential of pre-diagnostic mtDNAcn serving as a biomarker for predicting oxidative stress–related outcomes.

Additionally, as a reflection of oxidative stress levels, mtDNAcn may be especially closely associated with risks of obesity-related cancers, such as CRC (30). Overall obesity and abdominal adiposity may lead to increases in oxidative stress and systemic inflammation (32), and have been associated with elevated CRC risk (33). Our previous work showed that in healthy women, mtDNAcn was inversely associated with BMI even after adjusting for telomere length (TL) (18). Recently, Hang *et al.* also found that mtDNAcn tends to decrease continuously and persistently with adiposity over the life course (34). In addition, other environmental exposures such as exercise and smoking may also be involved in the regulation of mtDNAcn (35,36). For example, our own group found that duration and pack-years of smoking were inversely associated with mtDNAcn in leukocytes, while consumption of whole fruits and intake of flavanones (a group of antioxidants abundant in fruits) were positively associated with mtDNAcn (37). Smoking and physical inactivity are well-established risk factors for CRC (9,38), while fruit and vegetables may reduce risk (39). These prior data suggest that mtDNAcn could be a marker or mediator of the accumulating environmental exposures and associated systemic inflammation, and may exert an indirect influence on CRC risk.

To the best of our knowledge, this is the first prospective study examining the association

between mtDNAcn and CRC risk in a western population. Our study has several strengths, including its prospective design, long-term follow-up, pre-diagnostic assessment of mtDNAcn, and a comprehensive list of covariates. In addition, we include only incident CRC cases diagnosed after blood collection, which avoids the potential reverse influence of cancer progress and treatment effects on leukocyte mtDNAcn levels. In the sensitivity analysis, results barely changed after removing CRC cases diagnosed within 2 years after blood draw, suggesting that the observed association is unlikely the result of undiagnosed CRC present at blood draw.

Notably, in our study, we observed a significant inverse association for proximal colon cancer but not for distal colon or rectal cancer. This may be due to the small number of cancer cases at the latter sites. However, research has shown that clinical, pathological/histological, and molecular features differ between colon and rectal cancer, as well as between distal (left side) and proximal (right side) colon cancer (40-42). For example, proximal colon cancers are more likely to be microsatellite instability-high (MSI-high) tumors, while distal colon cancers are more likely to be chromosomal instability-high (CIN-high) tumors (42). Also, previous research has demonstrated that associations between environmental factors and CRC risk may be modified by tumor molecular subtypes (43-45). Recently, van Osch *et al.* found that, compared to other CRC tissues, mtDNAcn was significantly lower in CRC tissues with *BRAF* mutation (a mutated gene typically in MSI-high tumors) and those with high-level microsatellite instability (*MSI*), while mtDNAcn was higher in CRC tissues with *KRAS* mutation (a mutated gene typically in CIN-high tumors) (46). Whether these molecular features interact with mtDNAcn in modifying risk of CRC at subsites requires further investigation.

We acknowledged some limitations of our study. One is the relatively modest sample size in stratified subgroups, which limited the statistical power of interaction and stratified analyses.

Another limitation is the lack of detailed clinical-pathological characteristics and molecular classifications of these tumors. Future research investigating the mtDNAcn/CRC relationship by cancer molecular subtypes according to established markers and somatic profiles is needed.

In summary, in this nested case-control study, we found a significant inverse association between pre-diagnostic leukocyte mtDNAcn and CRC risk. Further investigations are warranted to explore whether mtDNAcn could become a valuable and long-term biomarker in evaluating the risk and prognosis of CRC. Importantly, additional basic experimental studies are needed to explore the biological mechanisms underlying the relationship between mtDNAcn and CRC carcinogenesis.

Supporting information:

1). Supplementary methods: mtDNAcn ascertainment and validation

Acknowledgements: This work is supported by NIH grants UM1 CA186107, R01 CA49449, and P01CA87969. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would like to thank the participants and staff of the Nurses' Health Study (NHS) for their valuable contributions, as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, and WY. The authors assume full responsibility for analyses and interpretation of these data.

Conflict of Interest Statement: None declared.

References

1. Siegel, R.L., *et al.* (2019) Cancer statistics, 2019. *CA Cancer J Clin*, **69**, 7-34.
2. Ferlay, J., *et al.* (2019) Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*, **144**, 1941-1953.
3. Schreuders, E.H., *et al.* (2015) Colorectal cancer screening: a global overview of existing programmes. *Gut*, **64**, 1637-49.
4. Fitzmaurice, C., *et al.* (2017) Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol*, **3**, 524-548.
5. Zole, E., *et al.* (2018) Mitochondria, its DNA and telomeres in ageing and human population. *Biogerontology*.
6. Larsen, N.B., *et al.* (2005) Nuclear and mitochondrial DNA repair: similar pathways? *Mitochondrion*, **5**, 89-108.
7. Stein, A., *et al.* (2017) Mitochondrial DNA repair and damage tolerance. *Front Biosci (Landmark Ed)*, **22**, 920-943.
8. Marley, A.R., *et al.* (2016) Epidemiology of colorectal cancer. *Int J Mol Epidemiol Genet*, **7**, 105-114.
9. Song, M., *et al.* (2018) Environmental factors, gut microbiota, and colorectal cancer prevention. *Clin Gastroenterol Hepatol*.
10. Reuter, S., *et al.* (2010) Oxidative stress, inflammation, and cancer: how are they linked? *Free Radic Biol Med*, **49**, 1603-16.
11. Carini, F., *et al.* (2017) Colorectal Carcinogenesis: Role of Oxidative Stress and Antioxidants. *Anticancer Res*, **37**, 4759-4766.
12. Qu, F., *et al.* (2011) Association between mitochondrial DNA content in leukocytes and colorectal cancer risk: a case-control analysis. *Cancer*, **117**, 3148-55.
13. Huang, B., *et al.* (2014) Association of leukocyte mitochondrial DNA copy number with colorectal cancer risk: Results from the Shanghai Women's Health Study. *Cancer Epidemiol Biomarkers Prev*, **23**, 2357-65.
14. Thyagarajan, B., *et al.* (2012) Mitochondrial copy number is associated with colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev*, **21**, 1574-81.
15. Nan, H., *et al.* (2013) Aspirin use, 8q24 single nucleotide polymorphism rs6983267, and colorectal cancer according to CTNNB1 alterations. *J Natl Cancer Inst*, **105**, 1852-61.
16. Meng, S., *et al.* (2016) Pre-diagnostic leukocyte mitochondrial DNA copy number and skin cancer risk. *Carcinogenesis*, **37**, 897-903.
17. Meng, S., *et al.* (2016) Pre-diagnostic leukocyte mitochondrial DNA copy number and risk of lung cancer. *Oncotarget*, **7**, 27307-12.
18. Meng, S., *et al.* (2016) Leukocyte mitochondrial DNA copy number, anthropometric indices, and weight change in US women. *Oncotarget*, **7**, 60676-60686.
19. Cao, Y., *et al.* (2016) Population-wide Impact of Long-term Use of Aspirin and the Risk for Cancer. *JAMA Oncol*, **2**, 762-9.
20. Ainsworth, B.E., *et al.* (1993) Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sports Exerc*, **25**, 71-80.
21. Chiuve, S.E., *et al.* (2012) Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr*, **142**, 1009-18.

22. Wolf, A.M., *et al.* (1994) Reproducibility and validity of a self-administered physical activity questionnaire. *Int J Epidemiol*, **23**, 991-9.
23. Willett, W.C., *et al.* (1987) Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. *J Am Diet Assoc*, **87**, 43-7.
24. Xie, H., *et al.* (2013) Reduced mitochondrial DNA copy number in peripheral blood leukocytes increases the risk of soft tissue sarcoma. *Carcinogenesis*, **34**, 1039-43.
25. Hofmann, J.N., *et al.* (2014) 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*, **35**, 1028-31.
26. Kim, C., *et al.* (2015) Mitochondrial DNA copy number and chronic lymphocytic leukemia/small lymphocytic lymphoma risk in two prospective studies. *Cancer Epidemiol Biomarkers Prev*, **24**, 148-53.
27. Lee, H.C., *et al.* (2005) Mitochondrial biogenesis and mitochondrial DNA maintenance of mammalian cells under oxidative stress. *Int J Biochem Cell Biol*, **37**, 822-34.
28. Shokolenko, I., *et al.* (2009) Oxidative stress induces degradation of mitochondrial DNA. *Nucleic Acids Res*, **37**, 2539-48.
29. Tyrka, A.R., *et al.* (2015) Association of telomere length and mitochondrial DNA copy number in a community sample of healthy adults. *Exp Gerontol*, **66**, 17-20.
30. Blasco, M.A. (2005) Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet*, **6**, 611-22.
31. von Zglinicki, T. (2002) Oxidative stress shortens telomeres. *Trends Biochem Sci*, **27**, 339-44.
32. Rani, V., *et al.* (2016) Oxidative stress and metabolic disorders: Pathogenesis and therapeutic strategies. *Life Sci*, **148**, 183-93.
33. Song, M., *et al.* (2016) Long-term status and change of body fat distribution, and risk of colorectal cancer: a prospective cohort study. *Int J Epidemiol*, **45**, 871-83.
34. Hang, D., *et al.* (2018) Longitudinal associations of lifetime adiposity with leukocyte telomere length and mitochondrial DNA copy number. *Eur J Epidemiol*, **33**, 485-495.
35. Menshikova, E.V., *et al.* (2006) Effects of exercise on mitochondrial content and function in aging human skeletal muscle. *J Gerontol A Biol Sci Med Sci*, **61**, 534-40.
36. Lee, H.C., *et al.* (1998) Aging- and smoking-associated alteration in the relative content of mitochondrial DNA in human lung. *FEBS Lett*, **441**, 292-6.
37. Wu, S., *et al.* (2019) Fruit and vegetable consumption, cigarette smoke, and leukocyte mitochondrial DNA copy number. *Am J Clin Nutr*, **109**, 424-432.
38. Giovannucci, E. (2001) An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev*, **10**, 725-31.
39. Song, M., *et al.* (2015) Nutrients, foods, and colorectal cancer prevention. *Gastroenterology*, **148**, 1244-60.e16.
40. Missiaglia, E., *et al.* (2014) Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features. *Ann Oncol*, **25**, 1995-2001.
41. Tamas, K., *et al.* (2015) Rectal and colon cancer: Not just a different anatomic site. *Cancer Treat Rev*, **41**, 671-9.
42. Baran, B., *et al.* (2018) Difference Between Left-Sided and Right-Sided Colorectal Cancer: A Focused Review of Literature. *Gastroenterology Res*, **11**, 264-273.

43. Inamura, K., *et al.* (2016) Prediagnosis Plasma Adiponectin in Relation to Colorectal Cancer Risk According to KRAS Mutation Status. *J Natl Cancer Inst*, **108**.
44. Liu, L., *et al.* (2017) Association Between Inflammatory Diet Pattern and Risk of Colorectal Carcinoma Subtypes Classified by Immune Responses to Tumor. *Gastroenterology*, **153**, 1517-1530.e14.
45. Mehta, R.S., *et al.* (2017) Dietary Patterns and Risk of Colorectal Cancer: Analysis by Tumor Location and Molecular Subtypes. *Gastroenterology*, **152**, 1944-1953.e1.
46. van Osch, F.H., *et al.* (2015) Mitochondrial DNA copy number in colorectal cancer: between tissue comparisons, clinicopathological characteristics and survival. *Carcinogenesis*, **36**, 1502-10.

Table 1. Basic characteristics of colorectal cancer cases and controls in the nested case-control study within the NHS

Characteristics	Cases (n=324)	Controls (n=658)
Age at blood draw, mean (SD)	58.9 (6.7)	59.3 (6.6)
Age at diagnosis, mean (SD)	67.4 (7.5)	-
Caucasians, %	98.2	99.7
log-mtDNAcn, mean (SD)	-0.1 (0.3)	0.01 (0.3)
Regular users of aspirin, %	38.0	47.0
Regular users of non-aspirin NSAIDs, %	12.7	19.8
BMI, kg/m2, mean (SD)	25.7 (5.0)	25.6 (4.7)
Weight change from blood collection until 2 years before diagnosis, kg, mean (SD)	1.0 (6.5)	1.8 (6.9)
Physical activity, MET-hours/wk, mean (SD)	18.1 (18.0)	18.6 (19.1)
AHEI score, mean (SD)	46.4 (9.2)	47.3 (9.6)
Smoking status, % - past-smokers	39.2	42.3
- current-smokers	17.3	12.2
Alcohol consumption, g/d, mean (SD)	7.3 (12.3)	6.9 (10.5)
CRC in a parent or sibling, %	17.3	15.5
Parity, % - 0 child	6.5	4.4
- 1 child	5.6	8.4
- 2 children	25.3	23.9
- 3+ children	62.0	62.3
Postmenopasual status, %	86.1	88.6
Current postmenopausal hormone use, %	34.0	43.7

Notes:

1. Abbreviation: BMI: Body Mass Index; AHEI: Alternate Healthy Eating Index; MET: metabolic equivalent; 2. Values are means (SD) for continuous variables and percentages for categorical variables; 3. Percentage of current postmenopausal hormone use is calculated among postmenopausal women.

Table 2. Age-standardized basic characteristics by mtDNAcn quartiles among controls in this nested case-control study within the NHS

Characteristics	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Number of participants	164	165	165	164
log-mtDNAcn, mean (SD)	-0.4 (0.1)	-0.1 (0.0)	0.1 (0.0)	0.5 (0.2)
Age at blood draw, mean (SD)*	59.4 (6.9)	59.4 (6.5)	59.2 (6.4)	59.0 (6.7)
Regular users of aspirin, %	53.1	47.3	49.4	48.9
Regular users of non-aspirin NSAIDs, %	19.4	18.5	13.9	23.3
BMI, kg/m^2, mean (SD)	25.6 (3.4)	25.7 (3.2)	26.1 (4.1)	25.7 (4.1)
Weight change from blood collection until 2 years before diagnosis, kg, mean (SD)	1.7 (5.1)	2.5 (6.6)	0.8 (4.3)	1.3 (4.2)
Physical activity, MET-hours/wk, mean (SD)	16.9 (12.4)	18.6 (12.9)	19.2 (15.3)	20.9 (17.8)
AHEI score, mean (SD)	48.2 (8.0)	48.3 (7.0)	48.6 (7.3)	47.3 (8.3)
Smoking status, % - past-smokers	46.1	40.0	44.8	43.2
- current-smokers	13.9	13.2	7.0	9.0
Alcohol consumption, g/d, mean (SD)	6.7 (8.2)	6.8 (7.8)	7.6 (9.0)	6.9 (8.6)
CRC in a parent or sibling, %	19.3	17.2	16.0	16.8
Parity (≥ 2 children), %	88.5	88.6	81.0	88.8
Postmenopausal women, %	90.1	92.1	91.2	93.4
Current postmenopausal hormone use, %	47.5	42.1	39.3	42.6

Notes:
1. Abbreviation: BMI: Body Mass Index; AHEI: Alternate Healthy Eating Index; MET: metabolic equivalent; 2. Values are means (SD) for continuous variables and percentages for categorical variables, and are standardized to the age distribution of the study population; 3. * Value is not age adjusted; 4. Percentage of current postmenopausal hormone use is calculated among postmenopausal women.

Table 3. Associations of mtDNAcn with the risk of overall colorectal cancer, as well as cancers at anatomic subsites

	4th quartile	3rd quartile OR (95%CI)	2nd quartile OR (95%CI)	1st quartile OR (95%CI)	*P for trend*
Colorectal cancer					
Cases/controls (324/658)	57/164	68/165	78/165	121/164	
Model 1	ref	1.22 (0.78, 1.90)	1.42 (0.93, 2.18)	2.19 (1.47, 3.27)	< 0.0001
Model 2	ref	1.10 (0.69, 1.76)	1.40 (0.89, 2.19)	2.19 (1.43, 3.35)	< 0.0001
Colon cancer					
Cases/controls (253/509)	44/132	51/119	62/128	96/130	
Model 1	ref	1.28 (0.77, 2.12)	1.50 (0.94, 2.42)	2.27 (1.45, 3.55)	0.0002
Model 2	ref	1.15 (0.66, 1.99)	1.53 (0.92, 2.55)	2.28 (1.40, 3.71)	0.0003
Proximal colon cancer					
Cases/controls (151/300)	22/76	36/72	30/78	63/74	
Model 1	ref	1.76 (0.89, 3.50)	1.50 (0.77, 2.91)	3.06 (1.67, 5.61)	0.0003
Model 2	ref	1.71 (0.80, 3.69)	1.70 (0.82, 3.51)	3.31 (1.70, 6.45)	0.0003
Distal colon cancer					
Cases/controls (90/185)	18/48	14/43	31/45	27/49	
Model 1	ref	0.87 (0.38, 2.00)	1.70 (0.81, 3.54)	1.26 (0.58, 2.74)	0.28
Model 2	ref	0.51 (0.16, 1.56)	1.20 (0.46, 3.15)	1.15 (0.42, 3.12)	0.47
Rectal cancer					
Cases/controls (71/149)	13/32	17/46	16/37	25/34	
Model 1	ref	1.01 (0.40, 2.58)	1.13 (0.43, 2.94)	1.90 (0.78, 4.59)	0.10
Model 2	ref	0.73 (0.23, 2.26)	0.94 (0.30, 2.92)	1.91 (0.67, 5.46)	0.12

Notes:

1. Abbreviation: ref: reference group; OR: odds ratio; CI: confidence interval;
2. Model 1: Conditional logistic regression model, no covariates adjustment;
3. Model 2: Conditional logistic regression model, adjusting for body mass index (in tertiles: 0-23.2 kg/m2, 23.2-26.6 kg/m2, ≥ 26.6 kg/m2), physical activity (in tertiles, 0-8.2, 8.2-20.2, ≥ 20.2 MET-hours/week), weight change from blood collection until 2 years before diagnosis (in tertiles, < 0kg, 0-2.72kg, ≥ 2.72kg), smoking status (never, former, or current smokers), alcohol consumption (in tertiles, 0-0.8 g, 0.8-5.8 g, ≥ 5.8 g per day), menopausal status and postmenopausal hormone use (premenopausal, non-current users, current users), parity (0/1/2/3+children), Alternate healthy eating index (AHEI) (in tertiles, 0-42.8, 42.8-51.3, ≥ 51.3), regular aspirin use (yes/no), regular non-aspirin NSAIDs use (yes/no), family history of colorectal cancer (yes/no).

Table 4. Associations between mtDNAcn and colorectal cancer risk by time of follow-up since blood collection

	4th quartile	3rd quartile OR (95%CI)	2nd quartile OR (95%CI)	1st quartile OR (95%CI)	*P for trend*
< 5 years (n= 330)					
Cases/controls (94/236)	18/56	19/70	18/64	39/46	
Model 1	ref	0.89 (0.40, 1.97)	0.94 (0.42, 2.08)	3.03 (1.39, 6.60)	0.002
Model 2	ref	0.80 (0.33, 1.90)	0.86 (0.36, 2.06)	3.44 (1.43, 8.27)	0.003
≥ 5 years (n= 652)					
Cases/controls (230/422)	39/108	49/95	60/101	82/118	
Model 1	ref	1.47 (0.86, 2.50)	1.76 (1.06, 2.93)	1.96 (1.23, 3.13)	0.004
Model 2	ref	1.48 (0.82, 2.66)	1.72 (0.98, 3.00)	1.98 (1.18, 3.33)	0.009
≥ 8 years (n= 492)					
Cases/controls (176/316)	34/91	35/66	43/72	64/87	
Model 1	ref	1.45 (0.81, 2.62)	1.67 (0.96, 2.93)	1.95 (1.17, 3.26)	0.01
Model 2	ref	1.58 (0.79, 3.17)	1.59 (0.82, 3.07)	2.09 (1.14, 3.83)	0.02
≥ 10 years (n=339)					
Cases/controls (123/216)	23/62	25/43	32/47	43/64	
Model 1	ref	1.56 (0.77, 3.15)	1.97 (1.00, 3.90)	1.76 (0.95, 3.24)	0.07
Model 2	ref	1.38 (0.58, 3.30)	2.13 (0.92, 4.92)	1.92 (0.94, 3.95)	0.06

Notes:

1. Abbreviation: ref: reference group; OR: odds ratio; CI: confidence interval;
2. Model 1: Conditional logistic regression model, no covariates adjustment;
3. Model 2: Conditional logistic regression model, adjusting for body mass index (in tertiles: 0-23.2 kg/m2, 23.2-26.6 kg/m2, ≥ 26.6 kg/m2), physical activity (in tertiles, 0-8.2, 8.2-20.2, ≥ 20.2 MET-hours/week), weight change from blood collection until 2 years before diagnosis (in tertiles, < 0kg, 0-2.72kg, > 2.72kg), smoking status (never, former, or current smokers), alcohol consumption (in tertiles, 0-0.8 g, 0.8-5.8 g, ≥ 5.8 g per day), menopausal status and postmenopausal hormone use (premenopausal, non-current users, current users), parity (0/1/2/3+children), Alternate healthy eating index (AHEI) (in tertiles, 0-42.8, 42.8-51.3, ≥ 51.3), regular aspirin use (yes/no), regular non-aspirin NSAIDs use (yes/no), family history of colorectal cancer (yes/no).