1.a. Full Title: Epigenome-wide association study of nut consumption

b. Abbreviated Title (Length 26 characters): EWAS of nuts

2. Writing Group:

Writing group members:

Carolina Ochoa-Rosales¹
Casey M. Rebholz²
Myriam Fornage³
Rui Xia³
Kim Braun¹
Trudy Voortman¹

1. Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands
2. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
3. The Brown Foundation Institute of Molecular Medicine for the Prevention of Human Diseases at the University of Texas Health Sciences Center at Houston, Houston, Texas, USA

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __COR___ [please confirm with your initials electronically or in writing]

First author: Carolina Ochoa-Rosales
Address: Office Na- 2901, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands

   Phone: +31 630 59 53 74
   E-mail: c.ochoarosales@erasmusmc.nl

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

   Name: Casey M. Rebholz
   Address: 2024 East Monument Street, Suite 2-500
   Baltimore, Maryland 21287

   Phone: 410-502-2359
3. **Timeline:** The data to be used for the proposed analysis already exist. Phenotypic dataset preparation will be completed by Casey Rebholz and shared with Myriam Fornage. Data analysis, to be conducted by Myriam Fornage and Rui Xua, will be conducted upon approval of the manuscript proposal. Results (not data), including CpG site, estimate, standard error, p-value, and sample size, will be shared with Carolina Ochoa-Rosales and Kim Braun for meta-analysis with other cohort studies. We anticipate that a manuscript draft will be prepared for ARIC Publications Committee within two years of proposal approval.

4. **Rationale:**

   Recent evidence from clinical trials and observational studies relates nuts consumption, especially walnuts intake, with reduced risk of type 2 diabetes (T2D), and improvement of glycemic traits profile in T2D patients and healthy subjects.\(^1\)-\(^4\) However, the molecular mechanism underlying this association is unclear.

   DNA methylation is one epigenetic mechanism that regulates gene expression, and can be modulated by dietary exposures.\(^5\),\(^6\) A randomized controlled trial found differentially methylated sites at the DNA (CpGs) after five years of a Mediterranean diet intervention including nuts, compared to a low-fat control diet.\(^7\) Also, changes in DNA methylation have been identified in T2D and insulin resistance pathophysiology.\(^8\)

   Therefore, our aim is to investigate whether nuts intake is associated with DNA methylation at genes are involved in glucose metabolism pathway and whether the methylation of the identified CpGs is associated with glycemic traits. Further, we aim to study whether the methylation pattern at the nuts intake-associated CpGs plays a role in modulating the risk of T2D.

5. **Main Hypothesis/Study Questions:**

   1. To find epigenetic modifications related to total nuts intake or to specific type of nuts (such as walnuts and peanuts), implicated in glucose homeostasis and T2D development.
   2. To study whether the nuts-related identified CpGs are associated with glycemic traits and T2D.
   3. To determine whether the methylation at the nuts-related CpGs mediates the effect of nuts intake on glycemic traits and T2D.

6. **Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).**

   **Design & Analysis for Objective #1 -** To find epigenetic modifications related to total nuts intake or to specific type of nuts (such as walnuts and peanuts), implicated in glucose homeostasis and T2D development.

   **Exposures:**
Four exposures of nuts intake to be analyzed (where available), measured as grams per day adjusted by daily energy intake (kcal/day). Please see attached syntax to perform the daily energy adjustment.

1. Total nuts (gr/day - energy adjusted)
2. Total nuts + seeds (gr/day - energy adjusted)
3. Walnuts (gr/day - energy adjusted)
4. Peanuts (gr/day - energy adjusted)
5. Peanut butter (gr/day - energy adjusted)

Outcome: DNA methylation using the Illumina Infinium HM450K BeadChip array. Please use untransformed beta values normalized using the Functional Normalization method. If this is not possible, other normalization methods are also acceptable.

Adjustment Covariates:
- Age
- Sex (1 = female; 0 = male)
- Kcal = daily kilocalories of intake
- Smoking history (current smoker = 1 ; former smoker = 2 ; never smoker = 3)
- Body mass index (BMI), measured as kg/m²
- Cardiometabolic disease (0=No; 1=Yes), such as prevalent type 2 diabetes and prevalent coronary heart disease
- Dyslipidemia (0=No; 1=Yes), measured as fasting low density lipoprotein >70 mg/dL and/or triglycerides > 150 mg/dL and/or the use of lipid lowering medication
- Diet quality score index, as available (e.g. adherence to dietary guidelines)
- Technical covariates:
  - White blood cell count (WBC): In the same blood sample from which DNA was extracted. If you did not measure white blood percentages in the same sample as used for the DNA methylation measurement, please estimate WBC percentages using a prediction method (e.g. Houseman’s method). For example, in the NFBC, we use the following WBC as covariates: B-cells, granulocytes, monocytes, NK, CD4+ T and CD8+ T cells.
  - Batch effects + Cohort Specific covariates: Please correct for batch effects covariates. For example Array-Number, Position-on-Array. Also, correct for other cohort-specific covariates as you deem necessary. If your cohort includes multiple population ancestries, please take into account adjusting for principal components, as you think is most appropriate for your cohort

Analysis: Linear Mixed-Effect Model (lmer function in R)

Models:

**Model 1** \( \rightarrow \) CpG methylation \( \sim \) exposure + CD8T + CD4T + NK + Bcell + Mono + Gran + (1|Array_Number) + (1|Position_on_Array) + age + sex + kcal/day + smoking status
**Model 2 →** CpG methylation ~ exposure + CD8T + CD4T + NK + Bcell + Mono + Gran + (1|Array_Number) + (1|Position_on_Array) + age + sex + kcal/day + smoking status + diet quality score + physical activity

**Model 3 →** CpG methylation ~ exposure + CD8T + CD4T + NK + Bcell + Mono + Gran + (1|Array_Number) + (1|Position_on_Array) + age + sex + kcal/day + smoking status + diet quality score + physical activity + body mass index + cardiometabolic disease + dyslipidemia

**Sensitivity analyses:** In addition to running the EWAS using the 3 models described above, also run the same models for women and men separately and for black and white participants separately.

**Design & Analysis for Objective #2 - To study whether the nuts-related identified CpGs are associated with glycemic traits and T2D.** The identified CpGs associated with nuts consumption will be tested in its relationship with glycemic traits and T2D status.

**Analysis:** Linear Mixed-Effect Model (lmer function in R) for both continuous and categorical outcomes when adjusting for random effects.

**Outcomes:**

i) glycemic traits: fasting glucose, fasting insulin and HOMA-IR levels

ii) prevalent and incident T2D, and prevalent and incident pre-diabetes

**Exposure:** DNA methylation using the Illumina Infinium HM450K BeadChip array. Please use untransformed beta values normalized using the Functional Normalization method. If this is not possible, other normalization methods are also acceptable.

**Models:**

1) Prevalent T2D ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + (1|Array_Number) + (1|Position_on_Array) + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia

2) Incident T2D ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + (1|Array_Number) + (1|Position_on_Array) + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia

3) Prevalent pre-T2D ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + (1|Array_Number) + (1|Position_on_Array) + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia

Prevalent T2D cases must be removed from this analysis.

4) Incident pre-T2D ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + (1|Array_Number) + (1|Position_on_Array) + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia

Prevalent T2D cases must be removed from this analysis.
5) Fasting glucose ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + (1|Array_Number) + (1|Position_on_Array) + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia
Prevalent T2D cases and non-fasting samples must be removed from this analysis.

6) Fasting insulin ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + (1|Array_Number) + (1|Position_on_Array) + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia
Prevalent T2D cases and non-fasting samples must be removed from this analysis.

7) HOMA-IR ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + (1|Array_Number) + (1|Position_on_Array) + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia
Prevalent T2D cases and non-fasting samples must be removed from this analysis.

Sensitivity analyses:
   i) same models, sex stratified and, separately, race-stratified.

Design & Analysis for Objective #3 - To determine whether the methylation at the nuts-related CpGs mediates the effect of nuts intake on glycemic traits and T2D. The significant associations between the nuts-related CpGs and glycemic traits and T2D status will be investigated in a causal mediation analysis.

Analysis: Linear regression for continuous variables and logistic regression for categorical variables, using the mediation package in R.

Outcomes:
   i) glycemic traits: fasting glucose, fasting insulin and HOMA-IR levels
   ii) prevalent and incident T2D, and prevalent and incident pre-diabetes

Exposure: DNA methylation using the Illumina Infinium HM450K BeadChip array. Please use untransformed beta values normalized using the Functional Normalization method. If this is not possible, other normalization methods are also acceptable.

Models:
   1) Prevalent T2D ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia
   2) Incident T2D ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia
   3) Prevalent pre-T2D ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia
Prevalent T2D cases must be removed from this analysis.

4) Incident pre-T2D ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia
   Prevalent T2D cases must be removed from this analysis.

5) Fasting glucose ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia
   Prevalent T2D cases and non-fasting samples must be removed from this analysis.

6) Fasting insulin ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia
   Prevalent T2D cases and non-fasting samples must be removed from this analysis.

7) HOMA-IR ~ identified CpG methylation + CD8T + CD4T + NK + Bcell + Mono + Gran + age + sex + smoking status + physical activity + body mass index + cardiometabolic disease + dyslipidemia
   Prevalent T2D cases and non-fasting samples must be removed from this analysis.

Sensitivity analyses:
   i) Same models, but excluding cell count variables: CD8T, CD4T, NK, Bcell, Mono and Gran. Same models, sex stratified and, separately, race-stratified.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes  __X__ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES_DNA = “CVD Research” would be used? ____ Yes  ____ No
   (This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? __X__ Yes  ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = “No use/storage DNA”? __X__ Yes  ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at:  http://www.cscc.unc.edu/aric/mantrack/maintain/search/dtSearch.html
10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

MS #2322: DNA methylation-related SNPs interact with fatty acids and triglycerides
MS# 3146: DNA methylation in relation to diet quality in the CHARGE consortium
MS# 3147: EWAS on coffee and tea consumption

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? ____ Yes ___X__ No

11.b. If yes, is the proposal

___ A. primarily the result of an ancillary study (list number* _________)
___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __________ __________ __________)

*ancillary studies are listed by number at https://www2.cscc.unc.edu/aric/approved-ancillary-studies

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is your responsibility to upload manuscripts to PubMed Central whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References: