1.a. Full Title: Circulating and tissue biomarkers of omega-3 PUFA and incident type 2 diabetes

b. Abbreviated Title (Length 26 characters): N-3 Biomarkers and T2D

2. Writing Group:
   Writing group members: This is a large consortium analysis. Preliminary list of participating studies and investigators are provided below. The list is subject to minor changes.
   ARIC: Lyn M. Steffen, Xia Zhou, Weihua Guan
   CHS: Heidi Lai, Rozenn Lemaitre
   CCCC: Kuo-Lion Chien
   CR-Adults and GOLDN: Stella Aslibekyan
   FHS and WHIMS: Nathan Tintle
   KIHDD: Jyrki Virtanen
   MCCS: Graham Giles, Julie Bassett
   MetSIM: Markku Laakso, Maria Lankinen
   MESA: Alexis Frazier-Wood, Brian Steffen
   PIVUS: Ulf Riserus, Lars Lind, Matti Marklund
   3C Study: Cecilia Samieiri
   ULSAM study 50, ULSAM study 70: Ulf Riserus, Lars Lind, Matti Marklund
   EPIC-InterAct: Nita Forouhi, Fumiaki Imamura
   IRAS-FS: Lynne Wagenknecht, Nicholette Allred
   AOC: Marianne Geleijnse
   AGESR: Rachel Murphy
   Hisayama: Toshiharu Ninomiya
   NHS and HPFS: Frank Qian, Andres Ardisson Korat, Qi Sun, Frank B. Hu
   Senior authors: Qi Sun, Dariush Mozaffarian

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

First author: Frank Qian
Address: 924 E 57th St, Suite 104, Chicago, IL 60637

Phone: (615) 738-1218 Fax:
E-mail: nhfrq@channing.harvard.edu
**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: Lyn M. Steffen  
Address: 1300 S 2nd St, Room 300 West Bank Office Building, Minneapolis, MN 55454  
Phone: (612) 625-9307  
Fax: (612) 624-0315  
E-mail: steff025@umn.edu

3. **Timeline**: Cohort-specific analysis will be completed by end of January 2018. Pooled analysis will be completed in late February or early March 2018. Preliminary manuscript will be completed in two months, to be circulated with all participating investigators by May 2018.

4. **Rationale**: The association between omega-3 (n-3) polyunsaturated fatty acids (PUFA), especially those from marine sources, and incident type 2 diabetes (T2D) has gained interest given their potential protective effects against cardiovascular disease (CVD) (1). n-3 PUFA are obtained from plant sources in the form of alpha linoleic acid (ALA) or eicosapentanoic acid (EPA), docosapentanoic acid (DPA), and docosahexanoic acid (DHA) from seafood sources. The association between fish intake and T2D risk has been explored extensively in prospective cohort studies, but with conflicting results (2; 3). Meta-analyses of these studies have suggested significant heterogeneity, including by region: fish intake is positively associated with T2D risk in cohorts in North America and Europe, and inversely with risk in cohorts in Asia (4). However, the validity and potential biological mechanisms underlying these observations, which are based on self-reported fish intake from questionnaires, are unknown.

In contrast, fewer studies have examined objective circulating biomarkers of n-3 PUFA (ALA, EPA, DPA, and DHA) in relation to incident T2D (3). Such biomarkers reduce challenges of biased recall and memory errors. But, existing studies do not permit strong conclusions, and potential for publication bias remains. Therefore, more studies are warranted to further elucidate the association between long-chain n-3 PUFA biomarkers and T2D risk, especially given the heterogeneity reported in studies for fish consumption. The importance of fish consumption in many populations, the increased availability of foods such as dairy products and eggs that are fortified with n-3 fatty acids, and the growing use of fish oil supplements all render the effect of n-3 PUFA on T2D an important scientific, clinical, and public health question (5; 6).

Several biological mechanisms suggest potential benefits of n-3 PUFA for preventing T2D. Meta-analysis of short-term RCTs of fish oil supplementation indicate potential reductions in adiposity including body weight and waist circumference, as well as increased adiponectin and reduced circulating triglycerides (7). On the other hand, while there was initial concern regarding the potential of fish oil supplementation to worsen glycemic control among patients with T2D (8), recent meta-analyses and reviews have not identified any significant effect on glycemic control (9; 10).

Importantly, possible heterogeneity among populations may exist for the effect of n-3 PUFA on T2D, which has not been examined in detail in prior studies. For example, studies have
suggested potential sex-differences in the relationship of n-3 PUFA with both adiposity and T2D (7; 12) (13; 14). Also, as noted above, self-reported fish intake has divergent associations with T2D in Asian vs. Western cohorts, a finding of unknown biologic significance which requires investigation and replication. Genetics could also play a role, for instance interaction between n-3 PUFA and known genetic variants that alter risk of T2D.

The limited statistical power in prior analyses to assess potential biological differences in lipid compartments of these fatty acids, the existence of possible effect modification by sex, race/ethnicity, world region, or genetics, as well as publication bias on both studies of fish intake and biomarkers of long-chain omega-3 fatty acids is an indication that further studies are warranted.

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

**Objective:** To examine the association between biomarkers of n-3 PUFA with incident T2D:

**Specific Aim 1:** To investigate whether circulating/adipose tissue ALA, EPA, DPA, DHA, and EPA + DPA + DHA are associated with risk of incident T2D.

**Specific Aim 2:** To investigate potential effect modification by age, sex, race/ethnicity, geographical location (North America, Europe, Asia/Oceania), and linoleic acid (18:2n-6) biomarker concentrations, and (in cohorts with available data) an allele-based genetic risk score of T2D.

**Hypotheses:**

We hypothesize that circulating and tissue levels of ALA, EPA, DPA, DHA, and the sum of EPA, DPA, and DHA are associated with lower risk of 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).**

**Methods:** For this project, a similar approach and protocol will be used as the previous one regarding n-3 PUFA and CHD risk. In brief, we intend to include all participants who are adults (≥ 18yrs) and not diagnosed at baseline with T2D.

Prospective cohort studies and nested case-control studies with available data on ALA, EPA, DPA, and DHA biomarkers and outcomes are eligible to participate in the project. Exposures to be assessed include all available biomarkers [e.g. plasma (total plasma, phospholipid, cholesterol esters, or triglyceride), serum, erythrocytes, adipose tissues, or other, all as % total fatty acids] of ALA, EPA, DPA, and DHA.

**Population:** Included participants should meet the following criteria:

1. Adults (≥ 18yrs)
2. No prevalent T2D at time of fatty acid measurement
**Exposures:** Exposures to be assessed include all available biomarkers [e.g. in different lipid compartments including plasma or erythrocyte phospholipids, cholesterol esters, triglyceride, total plasma or serum, or adipose tissue, all as % total fatty acids] of the n-3 PUFA:

1. Alpha-linolenic acid (ALA; 18:3n-3)
2. Eicosapentaenoic acid (EPA; 20:5n-3)
3. Docosapentaenoic acid (DPA; 22:5n-3)
4. Docosahexaenoic acid (DHA; 22:6n-3)
5. EPA + DPA + DHA

**Analysis Plan for Each Participating Study**

For the purposes of this analysis, the relationship of each fatty acid of interest with T2D will be assessed:

1. Continuously, % total fatty acids per interquintile range – standardized difference between the midpoint of the first and fifth quintiles (10th and 90th percentiles). For nested case-control studies, we will use the interquintile range of the control group
2. Using study-specific quintiles to assess potential nonlinear dose-response, both as indicator variables and using restricted cubic splines

Analyses will be conducted separately within each lipid compartment and also pooled overall. When more than one fatty acid biomarker is available in any single cohort (e.g., the study has data for both plasma phospholipid and cholesterol ester fatty acids), all data will be used in compartment-specific analyses; and for pooling overall results, we will prioritize the lipid fraction which may best reflect long-term dietary intake, namely in the following order: adipose > phospholipid (plasma/erythrocyte) > total plasma/serum > cholesterol ester > triglyceride.

**Outcome definition**

Type 2 diabetes, defined as:

1) Fasting glucose concentration ≥ 126 mg/dL (7.0 mmol/L)
2) 2-hour post oral glucose tolerance test glucose concentration ≥ 200 mg/dL (11.1 mmol/L)
3) New use of an insulin or oral hypoglycemic medication
4) Fasting or non-fasting HbA1c concentration ≥ 6.5% or
5) Otherwise as defined by the study.

**Main Cohort-Specific Analysis**

Baseline descriptive information will be collected. Additionally, we will assess the correlations of EPA, DPA, and DHA with each other.

**Statistical Models:** We will examine the association of each fatty acid of interest with incident T2D using results from three regression models that adjust for different sets of covariates:

**Model 1**

- Age (years)
- Sex (male, female)
- Field site (if necessary, using cohort-specific categories as dummy variables)
- Race (using cohort-specific categories with Whites as reference, as dummy variables)
• Education (<High school, High school graduate, College or higher, if unavailable, cohort-specific categories as dummy variables)
• Occupation (using cohort specific categories, as dummy variables)
• Physical activity (kcal/week, METS/week, or hours/day; or cohort-specific definitions)
• Smoking (never, former, current)
• Alcohol use (drinks or servings/day, g/day or ml/day)
• Prevalent hypertension, treated or self-reported (Yes/No)
• Prevalent dyslipidemia, treated or self-reported (Yes/No)
• Prevalent coronary heart disease, treated or self-reported (Yes/No)
• BMI (kg/m²)
• Waist circumference (cm)
• Trans fatty acid biomarker concentrations, sum of all available (e.g., total t-18:1 and t-18:2) (% total fatty acids)
• Linoleic acid (LA; 18:2n-6) biomarker concentrations (% total fatty acids)

Model 2
• Covariates in Model 1 plus:
  • Triglycerides (mg/dL or mmol/L, or otherwise assessed in each study)

Model 3
• Covariates in Model 2 plus:
  • Fish consumption (servings/day or g/day, or otherwise assessed in each study)

Missing data: Participants with missing data on n-3 PUFA exposure should be excluded. For missing covariate data, a missing indicator category should be used for categorical covariates. Missing continuous covariates should be handled per the usual practice of each cohort and study investigators, e.g. imputation or exclusion.

Statistical Analysis:
For prospective cohort studies, Cox proportional hazards models, with robust variance, will be used to estimate the hazard ratio for T2D. For studies that used a case-cohort design, weighted Cox regression models will be used for estimating hazard ratio. Follow-up time will be calculated from baseline (biomarker measurement) to date of T2D diagnosis, end of follow-up, loss to follow-up, or death, whichever occurred first.

For nested-case control studies, conditional logistic regression analyses will be used to estimate odds ratios as proxies of relative risks or rate ratios, after confirming the sampling strategy in each study. For studies with risk-set sampling, the odds ratios are equivalent to hazard ratios.

For each exposure-outcome analysis, the β interaction coefficient and its robust standard error for each exposure will be recorded.

Sensitivity analyses: two separate sensitivity analyses will be conducted, on Model 1 only:
1. Cases identified in first 2 years after biomarker sampling will be excluded to minimize effect of reverse causation due to pre-existing health condition

2. Participants will be censored at the first 6 years of follow-up to minimize exposure misclassification due to within-person variation over time.

**Heterogeneity:** To examine heterogeneity, stratified analyses will be conducted in each cohort, using the covariates as specified in Model 1 only. For the following subgroups, the $\beta$ coefficient for each n-3 PUFA (as continuous variable) and their robust standard error (SE) will be recorded for each specified strata:

1. Age (< 60 years, ≥ 60 years)
2. BMI (<30kg/m$^2$, ≥30kg/m$^2$)
3. Sex (male, female)
4. Race (Caucasian, major race #2, major race #3 etc.)
5. LA biomarker concentration (< or ≥ median value in each study)
6. Triglycerides (<150 mg/dL or ≥150 mg/dL)

**Genetic risk analysis:** An additional analysis examining the interaction of genetic risk of T2D with circulating n-3 PUFA will be assessed in studies with available genetic information.

1. A standard list of T2D susceptibility loci that is available for all studies will be compiled. Each study will then calculate a genetic risk score (GRS) for each participant by counting the number of risk alleles and individual carries (0, 1, or 2 for each SNP). Each SNP will be weighted according to its relative effect size as measured by beta coefficient derived from meta-analyses of genetic risk factors of T2D.
2. The weighted GRS and its relationship with risk of T2D will be assessed in each cohort, per 1-SD increment, with adjustments for age, sex, and principal components.
3. Interactions between biomarker concentrations of each n-3 fatty acid (by quintiles) will be assessed to examine possible modification of genetic risk by each n-3 fatty acid, per 1-SD increase in GRS. Meta-analysis will be used to pool the results by quintiles.

**Meta-Analysis:**
Cohort-specific HRs and ORs (continuous and within each quintile) will be pooled using inverse-variance weighted fixed-effects meta-analysis. Random-effects meta-analysis will also be performed for comparison. Heterogeneity will be assessed using the $I^2$ statistic.

For subgroup analyses, effect estimates in each study-specific stratum will be similarly pooled. Statistical significance of differences between subgroups of potential sources of heterogeneity will be explored using inverse-variance weighted meta-regression.

Spline analyses: To statistically test possible nonlinear associations between each of the n-3 PUFA and T2D risk, we will model each of the associations with multivariate meta-regression restricted cubic splines. This result will be compared to the pooled HR for each study-specific quintile, evaluated as an indicator variable against the lowest quintile as the reference. This
analysis will only be performed for fractions with sufficient numbers of cohort-measures (e.g. 5+), likely only the total plasma and phospholipid fractions due to the limited number of prior studies with available for cholesterol esters, triglycerides, and adipose tissue.

7.a. Will the data be used for non-CVD analysis in this manuscript? ___X__ Yes    ____ 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? ___X__ 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/search.php

   ___X___ Yes    _______ No

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)? Omega-6 biomarkers and risk of type 2 diabetes (Jason Wu et al)

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

11.b. If yes, is the proposal
   ___ A. primarily the result of an ancillary study (list number)* MN fatty acid data
   ___ 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 http://www.cscc.unc.edu/aric/forms/

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.csecc.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.