1.a. **Full Title:** Gene-dietary fat interaction on type 2 diabetes incidence: CHARGE Consortium

b. **Abbreviated Title (Length 26 characters):** GxFat interaction on T2D

2. **Writing Group:**
   ARIC Writing group members: Anne Justice, Kari North, Kristin Young, Liz Selvin and other investigators welcome
   CHARGE Research team: Jordi Merino, Jaeyoung Hong, Marta Guasch-Ferre, Caren E Smith, Toshiko Tanaka, Josee Dupuis, Frank B Hu and Jose C Florez.

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. AEJ

**Contact Author: Anne Justice**
Address: 137 East Franklin Street, Suite 306
Department of Epidemiology
University of North Carolina at Chapel Hill
Chapel Hill, NC 27514
Phone: 919.966.1403 Fax: 919.966.9800
E-mail: anne.justice@unc.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: **Kari North**
Address: 137 East Franklin Street, Suite 306
Department of Epidemiology
University of North Carolina at Chapel Hill
Chapel Hill, NC 27514
Phone: 919.966.2148 Fax: 919.966.9800
E-mail: kari_north@unc.edu

3. **Timeline:** 1 year
   - Study-level statistical analyses: July-August 2016
   - Interpretation and meta-analyses: September-December 2016
   - Manuscript preparation: January-May 2017
   - Manuscript submission: June 2017

4. **Rationale:** The discipline of nutritional genomics has developed with the aim of deepening knowledge on the interaction between dietary factors and genetic background on the modulation of both phenotypic traits and disease risk\(^1\). In the past decade, massively parallel
DNA genotyping and sequencing has opened the door to the identification of genetic variants that help explain the inherited basis of type 2 diabetes (T2D), a global public health crisis, fueled by obesity epidemics, that comprises a major, growing cause of morbidity and premature death. These studies are providing insights into the genetic architecture of T2D, but they explain only a small proportion of the heritability. A potential reason may arise from the interplay between genetics and environmental determinants, where quality and quantity of dietary fat have been demonstrated to play an important role.

Randomized clinical trials evaluating lifestyle interventions that include or focus on percentage fat in diets have shown beneficial effects on glycemic traits and T2D incidence after both a hypocaloric low-fat diet and a hypocaloric high-fat diet. Observed benefits from both high- and low-fat diets may result from differences in genetic backgrounds among included study participants, confluence with other T2D-related phenotypes, especially obesity, or the fact that dietary fat quality seems to be more important than total dietary fat intake. The finding that a high-fat diet may be beneficial for some individuals is contradictory to current public health guidelines that still hold to recommendations of fat-free and low-fat diets for all Americans. To better understand the relationship between dietary fat intake and T2D, we will investigate the association between the quality and quantity of dietary fat and the incidence of T2D according to T2D-genetic susceptibility.

Genetic studies may help to clarify the relationship between quality and quantity of dietary fat intake and incident T2D because they incorporate an instrument (genetic variation) that precedes the phenotype and is not affected by it. However, this approach does not rule out the possibility that genetic variants associated with T2D may also be correlated with other risk factors for T2D such as obesity. This limitation can be avoided by using the cross-phenotype meta-analyses method (CPMA). This approach allows for the dissection of causal influences for T2D risk independent of the risk conferred by obesity. Thus, the main objective of the present proposal is to develop a genetic risk score (GRS), accounting for summary-level genome-wide association studies (GWAS) data on multiple T2D-increasing risk genetic variants, to evaluate the role of dietary fat intake on T2D incidence. Additionally, we will apply gene set enrichment/pathway analyses to derive potential mechanistic insight of the research proposal.

5. Main Hypothesis/Study Questions:
Aim: To determine the role of dietary fat intake on T2D incidence according to different T2D genetic predisposition. Additionally, we will apply a gene set enrichment analyses to derive potential mechanistic insights.

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).

Study design: Meta-analysis of prospective observational study data taken from cohort follow-up visits in the CHARGE Consortium.
Inclusion:
- Adults ≥ 18 years of age
- European and African American Ancestry

Exclusion:
- Missing genotype data or low quality imputation. It will be considered missing genotype data if a particular participant does not have any genetic information. Each cohort will define low quality imputation according its previous experience (usually r^2>0.3).
- Baseline diabetes (For ARIC, we will consider visit 2 as baseline)
- Extreme energy intake and implausible dietary data. Each cohort will define extreme energy intake or non-reliable dietary data based on their previous experience.
- Missing dependent variables.
- History of previous cardiovascular events, including those with coronary heart disease, cerebrovascular disease and peripheral artery disease.

Primary outcome: incident T2D (Diabetes at visit 5 defined as a random plasma glucose concentrations >= 11.1mmol/L (200mg/dl), fasting glucose >=7mmol/l (126mg/dl), being on diabetes medications or self-reported diagnosis (e.g. same criteria we will use to exclude individuals with diabetes as visit 2/baseline, DIABTS24), using ARIC variable DIABTS54.

Genotype data:
- 1KG Phase 1 Imputed genotypes or Directly Genotyped data

Independent Dietary Variables:
1. Total dietary fat (i. Percentage of energy intake, evaluated continuously and ii. Low total fat vs high total fat (percentage of energy intake less than 20% vs more than 35%, reference category: 20-35%, categorical variable)). Dietary fat will be measured as a cumulative dietary fat intake averages across visit 2 and visit 3.
2. Subtypes of dietary fat
   a. MUFA (i. Percentage of energy intake, evaluated continuously and ii. MUFA intake lower than 10% vs more than 15%, reference category: 10-15%, categorical variable)
   b. PUFA (i. Percentage of energy intake, evaluated continuously and ii. PUFA intake lower than 6% vs more than 10%, reference category: 6-10%, categorical variable)
   c. SFA (i. Percentage of energy intake, evaluated continuously and ii. SFA intake lower than 7% vs more than 10%, reference category: 7-10%,categorical variable)
   d. Trans fatty acids; evaluated continuously.
   e. Omega-3 fatty acids intake, evaluated continuously and excluding supplementation

Covariates:
Models for T2D incidence will be adjusted for the following covariates:

1. Age
2. Sex
3. Parental family history of T2D (if available)
4. Dyslipidemia defined as total cholesterol (TC) > 240 mg/dl or low density lipoprotein-cholesterol (LDL-C) > 160 mg/dl or high density lipoproteins cholesterol (HDL-C) < 40 mg/dl (male) or < 50 mg/dl (female) or triglycerides (TGs) > 200 mg/dl. (measured at visit 2)
5. Hypertension at visit 2 (systolic blood pressure > 140, and/or diastolic blood pressure > 90, and/or taking antihypertensive medication, HYPERT25).
6. BMI (kg/m2)
7. Physical activity (METs/h/wk) (measured at visit 3)
8. Smoking status. Defined as current, never or former.
9. Total energy intake (kcal/day) (cumulative average)
10. Nutrients (cumulative average of total protein intake (% energy), fiber intake (g/d), alcohol intake (g/d))

Summary data analysis:

1.1 Data sources and candidate instrument selection

To create an unbiased genetic instrument for T2D incidence, we identified genome-wide associated variants with T2D in the largest GWAS meta-analysis of T2D by the Diabetes Genetics Replication and Meta-analyses (DIAGRAM) Consortium study (34,840 cases / 114,981 controls). We pruned the list of T2D-related variants for linkage disequilibrium (LD) using SNAP (https://www.broadinstitute.org/mpg/snap/ldsearch.php) applied to HapMap European samples (n=68 SNPs). To avoid pleiotropy of T2D-increasing risk genetic variants with obesity, we interrogated each T2D-increasing risk genetic variant for potential overlap with obesity increasing risk variants in the largest GWAS meta-analyses for body mass index (BMI) including data from 339,224 individuals and waist circumference in up to 224,459 individuals using the cross-phenotype meta-analysis method (n=43 SNPs). We will harmonize the polygenic risk score (GRS) among cohorts and we will create an aggregate weighted GRS for T2D prediction based on the assumption of an additive genetic effect, by assigning 1 point for each risk allele (low-risk homozygotes = 0 points; heterozygotes = 1 point; high-risk homozygotes = 2 points). Each study participant will be assigned a quantitative GRS based on the number of risk alleles and their β-estimates present at the SNPs under investigation (see supplementary GRS formula). To build the GRS we will use successfully genotyped variants from each participating cohort with higher call rate than 0.95 and Hardy-Weinberg equilibrium higher than 1×10^-4. When not directly genotyped, variants will be imputed using 1000 Genomes Project according to quality imputation standards defined by each cohort.

Once the polygenic risk score is “conceptually” designed, each study analyst within CHARGE will receive a subset file composed of the included T2D-increasing risk variants, the effect allele for each variant and the β-estimate coefficients and standard error for each particular genetic variant. (Supplementary table 1). Then the analyst will create an aggregate weighted GRS for T2D prediction based on the assumption of an additive genetic effect, by assigning 1 point for
each risk allele (low-risk homozygotes = 0 points; heterozygotes = 1 point; high-risk homozygotes = 2 points). In doing so, each analyst will multiply the number of risk alleles present per SNP by the β-estimate reported for that SNP in the DIAGRAM Consortium GWAS\(^5\). Each study participant will be assigned a quantitative GRS based on the number of risk alleles and their β-estimates present at the SNPs under investigation. In building the GRS the analyst will use successfully genotyped variants from each participating cohort with higher call rate than 0.95 and Hardy-Weinberg equilibrium higher than 1×10\(^{-4}\). When not directly genotyped, variants will be imputed using 1000 Genomes Project according to quality imputation standards defined by each cohort.

**Analyses.** The analysis plan will be divided in 3 main steps (Table 1). First, multivariable Cox Regression models will be used to prospectively estimate the effect of the exposure (defined as a cumulative averages of dietary fat / subtypes) on incident type 2 diabetes expressed as hazard ratios (HR) and 95 confidence interval (95% CI). (Example of model: \textit{Incident T2D} \sim \textit{total dietary fat + covariates} (1-2). Second, we will re-run the same Cox Regression models including an interaction term (Example of model: \textit{Incident T2D} \sim \textit{total fat + GRS + total fat*GRS + covariates} (1-10). A detailed explanation of model’s exposures and analysis covariates is provided in Table 1. The last part of the analysis plan is to conduct a pathway analysis.

Table 1. Analysis outline.

<table>
<thead>
<tr>
<th>Outcome; Incident T2D</th>
<th>Exposure</th>
<th>Association Analysis (AA) covariates</th>
<th>Interaction Analysis (Exposure variable + GRS + Interaction term + AA covariates)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>Total dietary fat, cont.</td>
<td>Age, gender</td>
<td>N/A</td>
</tr>
<tr>
<td>Model 2</td>
<td>Total dietary fat, cont.</td>
<td>(1-10)</td>
<td>Total fat + GRS + Total fat*GRS +</td>
</tr>
<tr>
<td>Model 3</td>
<td>SFA, cont.</td>
<td>Age, gender</td>
<td>N/A</td>
</tr>
<tr>
<td>Model 4</td>
<td>SFA, cont.</td>
<td>(1-10) + MUFA + PUFA + trans</td>
<td>SFA + GRS + SFA*GRS +</td>
</tr>
<tr>
<td>Model 5</td>
<td>PUFA, cont.</td>
<td>Age, gender</td>
<td>N/A</td>
</tr>
<tr>
<td>Model 6</td>
<td>PUFA, cont.</td>
<td>(1-10) + MUFA + SFA + trans</td>
<td>PUFA + GRS + PUFA*GRS</td>
</tr>
<tr>
<td>Model 7</td>
<td>MUFA, cont.</td>
<td>Age, gender</td>
<td>N/A</td>
</tr>
<tr>
<td>Model 8</td>
<td>MUFA, cont.</td>
<td>(1-10) + PUFA + SFA + trans</td>
<td>MUFA + GRS + MUFA*GRS</td>
</tr>
<tr>
<td>Model 9</td>
<td>Trans, cont.</td>
<td>Age, gender</td>
<td>N/A</td>
</tr>
<tr>
<td>Model 10</td>
<td>Trans, cont.</td>
<td>(1-10) + PUFA + SFA + MUFA</td>
<td>Trans + GRS + Trans*GRS +</td>
</tr>
<tr>
<td>Model 11</td>
<td>Omega3, cont.</td>
<td>Age, gender</td>
<td>N/A</td>
</tr>
<tr>
<td>Model 12</td>
<td>Omega3, cont.</td>
<td>(1-10) + SFA + MUFA + trans</td>
<td>omega3 + GRS + omega3*GRS +</td>
</tr>
<tr>
<td>Model</td>
<td>Total dietary fat, cat.</td>
<td>Age, gender</td>
<td>N/A</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------</td>
<td>-------------</td>
<td>-----</td>
</tr>
<tr>
<td>Model 1.1</td>
<td>Total dietary fat, cat.</td>
<td>(1-10)</td>
<td>Total fat + GRS + Total fat*GRS +</td>
</tr>
<tr>
<td>Model 2.1</td>
<td>SFA, cat.</td>
<td>Age, gender</td>
<td>N/A</td>
</tr>
<tr>
<td>Model 3.1</td>
<td>SFA, cat.</td>
<td>(1-10)</td>
<td>SFA + GRS + SFA*GRS +</td>
</tr>
<tr>
<td>Model 4.1</td>
<td>MUFA, cat.</td>
<td>Age, gender</td>
<td>N/A</td>
</tr>
<tr>
<td>Model 5.1</td>
<td>MUFA, cat.</td>
<td>(1-10) + PUFA + MUFA + trans</td>
<td>PUFA + GRS + PUFA*GRS +</td>
</tr>
<tr>
<td>Model 6.1</td>
<td>PUFA, cat.</td>
<td>(1-10) + MUFA + SFA + trans</td>
<td>MUFA+ GRS + MUFA*GRS +</td>
</tr>
<tr>
<td>Model 7.1</td>
<td>PUFA, cat.</td>
<td>(1-10) + MUFA + SFA + trans</td>
<td>MUFA+ GRS + MUFA*GRS +</td>
</tr>
<tr>
<td>Model 8.1</td>
<td>MUFA, cat.</td>
<td>(1-10)</td>
<td>MUFA+ GRS + MUFA*GRS +</td>
</tr>
</tbody>
</table>

**Table legend:** cont; continuous (% energy intake), cat; categorical (low fat vs high fat), (See dictionary data).

**Meta-Analysis:** Summary statistics from each cohort derived by the interaction analyses will be combined using inverse variance-weighted, fixed-effects meta-analysis by the research team. Genetic heterogeneity in the meta-analyses will be quantified with the Higgins’ $I^2$ parameter. As a sensitivity analysis, the research team will perform a random-effects meta-analysis in which the weights assigned to individual estimates are adjusted to intrinsic variability. Also we will account for other sources of heterogeneity including age ranges or gender.

**Pathway analyses:** To assess mechanistic insights from the GRS we will use the Data-driven Expression Prioritized Integration for Complex Traits (DEPICT), a computational integrative tool based on predicted gene functions\(^{18}\). This method systematically prioritizes the most likely causal genes at associated loci and highlights enriched pathways. For the present proposal we will prioritize genes and regulatory elements to derive potential mechanistic insights.

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? **Yes** __X__ 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.csc.c.unc.edu/ARIC/search.php

_____ Yes __X____ 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)?

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 http://www.csc.c.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.csc.c.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.

13. Per Data Use Agreement Addendum for the Use of Linked ARIC CMS Data, approved manuscripts using linked ARIC CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu.

I will be using CMS data in my manuscript _____ Yes __X__ No.

Bibliography