1.a. Full Title: Genome-wide interaction with dietary fat intake with respect to fasting plasma LDL, HDL, and triglyceride concentrations
   b. Abbreviated Title (Length 26 characters): GWAS x dietary PUFA

2. Writing Group: the lead authorship positions will be shared, but the primary is from the Framingham cohort—information listed below: Julius S. Ngwa (pre-doctoral fellow with Adrienne Cupples)

   Writing group members: ARIC authors are listed below (others are welcome); in lieu of a full author list (TBD), other participating cohorts (from the CHARGE nutrition working group (WG)) are given in the table.

   Jenifer A. Nettleton (senior/ WG chair), Kari E. North (collaborator/ WG member) + others welcome

<table>
<thead>
<tr>
<th>Euro Am (n's approx)</th>
<th>Cohort rep and/or analyst</th>
<th>GWAS/ diet</th>
<th>METABOCCHIP/ diet</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIC</td>
<td>Nettleton</td>
<td>8600</td>
<td>NO</td>
</tr>
<tr>
<td>CHS</td>
<td>Lemaitre, Siscovick</td>
<td>3225</td>
<td>NO</td>
</tr>
<tr>
<td>Family HS</td>
<td>Wojczynski, Borecki</td>
<td>3644</td>
<td>NO</td>
</tr>
<tr>
<td>Framingham</td>
<td>Ngwa, Cupples, Ordovas</td>
<td>6010</td>
<td>NO</td>
</tr>
<tr>
<td>GLACIER</td>
<td>Renstrom, Franks</td>
<td>pending</td>
<td>6000</td>
</tr>
<tr>
<td>GOLDN</td>
<td>Frazier-Wood, Arnett</td>
<td>900</td>
<td>NO</td>
</tr>
<tr>
<td>HBCS</td>
<td>Ericksson</td>
<td>1584</td>
<td>NO</td>
</tr>
<tr>
<td>Health ABC</td>
<td>Houston, Harris</td>
<td>1400</td>
<td>NO</td>
</tr>
<tr>
<td>HPFS</td>
<td>Qi</td>
<td>1100</td>
<td>NO</td>
</tr>
<tr>
<td>InCHIANTI</td>
<td>Tanaka, Ferrucci</td>
<td>1185</td>
<td>NO</td>
</tr>
<tr>
<td>Malmo</td>
<td>Sonestedt, Orho-Melander</td>
<td>pending</td>
<td>2100</td>
</tr>
<tr>
<td>MESA</td>
<td>Nettleton, Rotter</td>
<td>2300</td>
<td>NO</td>
</tr>
<tr>
<td>NHS</td>
<td>Qi</td>
<td>1280</td>
<td>NO</td>
</tr>
<tr>
<td>PIVUS</td>
<td>Ingelsson</td>
<td>NO</td>
<td>844</td>
</tr>
<tr>
<td>Rotterdam</td>
<td>van Rooij, Zilikens</td>
<td>2560</td>
<td>NO</td>
</tr>
<tr>
<td>THISEAS</td>
<td>Dedoussis</td>
<td>500</td>
<td>725</td>
</tr>
<tr>
<td>ULSAM</td>
<td>Ingelsson</td>
<td>NO</td>
<td>1084</td>
</tr>
<tr>
<td>YFS</td>
<td>Lehtimäki, Mikkila</td>
<td>1775</td>
<td>NO</td>
</tr>
<tr>
<td>DILGOM</td>
<td>Salomaa</td>
<td>814</td>
<td>3952</td>
</tr>
<tr>
<td>Health 2000</td>
<td>Knekt</td>
<td>2124</td>
<td>NO</td>
</tr>
</tbody>
</table>

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

- **Cohort-specific data analyses (pilot):** March 30, 2012
- **Meta-analysis (pilot):** May 1, 2012 *(to be presented at CHARGE meeting in Iceland)*
- **Cohort-specific data analyses (full effort):** July 15, 2012
- **Meta-analysis (full effort):** August 30, 2012
- **Manuscript drafting complete:** November 1, 2012

4. **Rationale:**

Genetic and dietary factors are important determinants of an individual’s plasma lipid profile. Dietary recommendations to manage plasma lipid concentrations have focused on macronutrient composition, particularly types of major fats. However, focusing exclusively on the dietary fatty acid composition of a food without regard for other attributes might lead to inaccurate conclusions. Thus, more recently, greater attention has also been given to the overall context and quality of the diet, which includes a vast number of foods, their biologically active constituents and their synergic interactions. Concurrent to these subtle shifts in focus in the field of nutrition are genetic studies that have uncovered over 100 loci related to risk of cardiovascular disease (CVD) and levels of CVD risk factors, like plasma lipids and lipoproteins. Both nutrition and genetic research would benefit from an in-depth, sufficiently powered exploration of the interaction between these dietary and genetic factors. Moreover, the identification of interactions between genetic loci and diet may help to elucidate the molecular pathways that link them with dyslipidemia and, ultimately, cardiovascular disease (CVD) as well as inform the design of public health preventive interventions and translation to clinical medicine. Therefore we propose the following:

**Aim:** Interaction DISCOVERY: Via meta-analysis of data from CHARGE cohorts with genome-wide arrays, including Cardio-MetaboChip, we will test interactions between genotype and i) specific dietary fatty acids, ii) key animal- and plant-food sources of fatty acids, iii) dietary patterns, with respect to plasma lipid traits
5. Main Hypothesis/Study Questions:

The overall aim, provided above, is a component of our recently submitted DARING grant application (Discovery And Replication of Interactions between Nutrition and Genetics). In order to obtain pilot data that can be incorporated into a revision/re-application of DIGITAL/DARING (and presented at the poster session of the next CHARGE face-to-face meeting in Reykjavik, Iceland), we propose the following more narrowly-focused analysis plan to address our ‘pilot aim.’ The full analyses will be pursued soon after. The deadline for data-sharing for this pilot effort is March 30, 2012.

The Pilot effort (including only cohorts able to meet 3/30 deadline) will focus on only dietary PUFA intake. Later efforts (which will include the full set of cohorts) will be informed by these findings, as well as will incorporate other dietary exposures: SFA, MUFA, and subtypes of PUFA (n-3 and n-6), the CHARGE diet score and component food groups (related proposals: ARIC ms 1738 & 1807). It is likely that these various sub-aims will take form in multiple manuscripts, but it is difficult to know the exact number or arrangement until we know how many, if any, interactions are identified. As the process unfolds, we will update the records so that manuscript proposals match to each of the to-be-published efforts.

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: Pilot Analysis (Full-Effort, details above & below)

Dependent Variables

1. HDL Cholesterol (mmol/L)
2. LDL Cholesterol (mmol/L)
3. Triglycerides (mmol/L)—after natural log transformation

Exclusions

- Implausible dietary data (by cohort-specific definition)
- Non-white race
- Missing genotype data
- Missing plasma lipid measurements
- Taking lipid-lowering medications (after the pilot effort is complete, we may revisit this exclusion; changing from exclusionary criteria to covariate... TBD)

Independent Dietary Variables

Pilot: Polyunsaturated Fat (% of total calories consumed), considered continuously
Later effort:
Saturated Fat (% of total calories consumed), considered continuously
Monounsaturated Fat (% of total calories consumed), considered continuously
Healthy Diet Score & Score Food Group Components (same definition as used for fasting glucose & adiposity projects, ms numbers given above)

‘HEALTHY’ FOOD GROUPS (quartile ranks summed to generate diet score: Qt1 = 0 pts; Qt2 = 1 pt; Qt3 = 2 pts; Qt4 = 3 pts)
- Whole Grains (defined as in our previous project)
- Fruit (not including juice*— in all cohorts where this is feasible)
- Vegetables (not including white potatoes; not including legumes*)
- Fish (fin fish [no shellfish], not fried—in all cohorts where fried can be distinguished from baked, broiled, raw, etc.)
- Nuts (including peanuts and nut butters*)

‘UNHEALTHY’ FOOD/BEVERAGE GROUPS (quartile ranks reversed: Qt4 = 0 pts; Qt3 = 1 pt; Qt2 = 2 pts; Qt1 = 3 pts & then summed to generate diet score)
- Red Meat & Processed Meat (combined group*)
- Sugar-sweetened beverages (soda pop and sugar-sweetened, artificially fruit flavored juices)
- Fried Potatoes* (in cohorts where intake was quantified)
- Desserts & Sweets

Score theoretical maximum = 27; Score minimum = 0

Additional analyses will also consider the fatty acids in tertiles as follows. Calculate the tertiles for each fatty acid (saturated, poly- and mono-unsaturated fats). Create dichotomized variables for 1) saturated fat of lowest tertile versus other tertiles (coded 1= lowest, 0 = middle, highest), 2) polyunsaturated fat of highest tertile versus others (coded 1= highest, 0=middle, lowest), 3) monounsaturated fat of highest tertile versus others (coded 1= highest, 0=middle, lowest).

Genetic Variables
Genome-wide (via GWAS or MetaboChip scans) including the X chromosome, if possible, for genotyped or imputed SNPs coded as dosages. Genotyped SNPs will take on the values 0, 1, or 2 while imputed SNPs will be the expected dosage as output from imputation software being any number between 0 and 2. Imputation should be from Phase II HapMap.

Covariates
1. Primary analyses will adjust for the following:
   a. Age
   b. Sex
   c. Total energy consumed (calories)
   d. field center (as needed)
   e. Principal Components to adjust for population substructure, as needed

2. Secondary analyses will further adjust for the following (where PUFA is exposure of interest-Pilot effort):
   a. protein intake (% of energy)
   b. MUFA intake (% of energy)
   c. SFA intake (% of energy)

ANALYSES
Conduct genome-wide linear regression analyses or linear mixed effects models for family data with a lipid measure as the outcome. The primary independent variables of interest will be the main effect of each SNP and its interaction with a dietary measure. Conduct analyses so that you will output not only these regression parameters, their standard errors and p values, but
also their covariance estimate. The covariance estimate will be used to conduct 2 df tests of the
association of a SNP, taking possible interaction into account. As a first pass, include the
primary covariates in the model. The Model can be stated as follows:

MODEL: LIPID TRAIT = SNP + Diet trait + SNP*Diet trait + covariates

Genome-wide significance will be results with p value < 5 x 10^{-8}. Each study will conduct
analyses of its data and results will be meta-analyzed with fixed effects inverse variance
weighting of the regression coefficients.

Obtain robust standard errors and covariance estimate between the SNP beta and the
Interaction beta.

Software that can be used to obtain robust standard error and robust covariance estimates
1. QUICKTEST version 0.95 or newer (http://toby.freeshell.org/software/quicktest.shtml)
2. ProbABEL version 0.1-3 or newer (http://mga.bionet.nsc.ru/~yurii/ABEL/)
3. R – the gee.test function in geepack in R can be used to obtain robust estimates of the
   variance of a linear model (geepack appears to be a bit more stable than gee)

DATA SHARING

File naming scheme
Please use the following file naming scheme for the uploaded files.
Results file: STUDY.PHENOTYPE.DIET.MODEL.DATE.txt
Where:
STUDY: is a short (14 characters or less) identifier for the population studied.
PHENOTYPE has options: HDL, LDL, TG
DIET: options –POLY, SFA, MUFA, FOODS, DIETSC
MODEL: PRIM for primary model; SEC for secondary model that includes protein, MUFA, SFA
DATE: is the date on which the file was uploaded, in the format “YYYYMMDD”
Examples:
    ARIC.HDL.SAT.CONT.20120216.txt
    FRAMINGHAM.LDL.POLY.DICH.20120216.txt
    CHS.TC.DIETSC.CONT.20100101.txt
Please also upload an additional information file:
    STUDY.PHENOTYPE.DIET.MODEL.DATE.xls which will be provided to all cohorts as a pre-
structured form to be filled.
7.a. Will the data be used for non-CVD analysis in this manuscript?
   *Fasting LDL, HDL, and triglyceride concentrations are the primary outcomes*

   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*
   (This file ICTDER03 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*

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”?  *Yes*

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](http://www.cscc.unc.edu/ARIC/search.php)  *Yes*

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)?

   **Examples of others from the CHARGE Nutrition working group:**
   1534 “Interactions between whole grain intake and genotype with respect to fasting glucose concentrations in multiple cohorts within the CHARGE & MAGIC consortia”
   1577 “Interactions between zinc intake and SNPs and their impact on fasting blood glucose levels in multiple cohorts within the CHARGE and MAGIC consortia”
   1675 “Low density lipoprotein receptor related protein 1, fatty acids and anthropometric traits”
   1656 “Genome-wide association analysis of macronutrient intake”
   1738 “Interaction between a multi-factorial diet score and genetic loci for fasting glucose and insulin”
   1779 “Meta-analysis: FTO and MC4R genes, Dietary Intakes and Obesity”
   1803 “Interactions between dietary fat and lipoprotein lipase variants for plasma lipid and lipoprotein concentrations”
   1807 “Interaction between a multi-factorial diet score and genetic loci for body mass index (WHR) and waist:hip ratio (WHR)”

11. a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data?  *Yes*
   - GWAS via STAMPEDE & GENEVA, #2006.03
   - Interactions between Diet and Genes Related to Risk of Type II Diabetes, #2007.12
11.b. If yes—is the proposal a primarily the result of an ancillary study
(numbers 2007.12, 2006.03, 2010.11)
ARI C is one of at least 20 cohort studies contributing data to the CHARGE -based meta-
analysis. Since this work is a product of CHARGE which utilizes GWA data, ancillaries
related to STAMPEDE & GENVA are also acknowledged.

12. 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.

The lead author is aware of, and will comply with, this stipulation.