ARIC Manuscript Proposal # 1787

PC Reviewed: 5/10/11 **Status: A** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: **Meta-analysis of Genome-Wide Association Studies for intake of fish and dietary eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA)**

b. Abbreviated Title (Length 26 characters): *GWAS for Dietary Fish and total EPA+DHA*

2. Writing Group:

ARIC writing group members**:**
Jennifer Nettleton, Ph.D., University of Texas Health Science Center, School of Public Health (ARIC author and Convener of the CHARGE Nutrition working group)
Data Analyst = staff/student to be identified, @University of Texas Health Science Center, School of Public Health
Lyn Steffen, Ph.D., University of Minnesota of Public Health (ARIC author)
***Other ARIC investigators interested in contributing to this work are invited.*

Manuscript Lead Author
Dariush Mozaffarian, Ph.D., Harvard School of Public Health (Cardiovascular Health Study contributor and lead/senior author of manuscript)

**Other authors members TBD *(this is a multi-cohort effort—ARIC is one of at least 15 participating cohorts, including other cohorts from the CHARGE Nutrition Working Group—CHS, Framingham, Health ABC, Family Heart Study, MESA, InCHIANTI, Rotterdam, AGES, GLACIER, Nurses' Health Study, Health Professionals' Follow-up Study, Young Finns Study, ULSAM, THISEAS)*

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

First author: **Dariush Mozaffarian, M.D., Dr.P.H.**
Address: **Department of Epidemiology**
655 Huntington Avenue
Boston, Massachusetts 02115
E-mail: ***dmozaffa@hsph.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: **Jennifer Nettleton, Ph.D.**
Address: **E-641 RAS; 1200 Herman-Pressler; Houston, TX 77030**
Phone: **713-500-9367** Fax: **713-500-9264**
E-mail: ***jennifer.a.nettleton@uth.tmc.edu***

3. Timeline:
April-May 2011: Assemble cohorts, submit proposals
June-July 2011: Complete GWAS
July-Aug 2011: Meta-analysis, secondary analyses
Sep-Oct 2011: First manuscript drafting
Nov-Dec 2011: Revisions
Jan 2012: Submit manuscript

4. Rationale:

Consumption of finfish and shellfish (hereafter referred to as fish) is linked to lower risk of several chronic diseases, in particular lower risk of fatal coronary heart disease.[1-3] These beneficial associations in observational studies are supported by randomized controlled trials demonstrating favorable effects of fish or fish oil on numerous chronic disease risk factors.[1, 4-7]

In comparison to many other foods, there is quite broad variation in levels of fish consumption in most populations. In many Western countries, for example, approximately one-third of individuals consume no fish at all, approximately one-third consume fish but relatively rarely (less than once per week), and approximately one-third consume fish more frequently. Many personal and environmental differences likely contribute to this diversity in frequency of fish consumption, including differences in geographic residence (coastal vs. inland), foods consumed as a child, socioeconomic status, and so forth. The extent to which common genetic variation also contributes is not well established and may be variable for different foods.[8] A recent heritability study evaluating Danish dizygotic and monozygotic twins estimated that heritability of fish consumption was 17% in men and 61% in women (additive effects).[9] It is plausible, for example, that genes related to taste, digestion, fatty acid metabolism, or other unknown processes related to food preferences could influence a person's frequency of fish consumption. One could attempt to evaluate sets of candidate genes, but such investigation would be strongly limited by imperfect knowledge of which genes affect established systems/processes related to food preferences and, even more so, which genes affect currently unknown systems/processes related to food preferences.

As has been seen with many other phenotypes including physiologic risk factors, genome-wide (non-hypothesis-driven) approaches may lead to discovery of novel genes and biologic pathways related to the phenotype of interest. Yet, virtually no genome-wide analyses have been performed evaluating consumption levels of specific foods as the phenotype. In part, this may have been due to limited sample sizes of individual cohorts having both genetic and dietary information. The CHARGE Nutrition Working Group provides an excellent opportunity to determine, in large consortium, whether common genetic variation relates to frequency of fish consumption, a phenotype linked to risk of many chronic diseases.

References listed on page 6

5. Main Hypothesis/Study Questions:

Three GWAS will be conducted within each participating cohort and then meta-analyzed. The three outcomes of interest are detailed below

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).**

1. **Outcomes:** Estimated dietary consumption of:
    a. Total fish, including shellfish and fried fish.
        i. Units: Servings/day (sum the following)

'DTIA34' = canned tuna fish
'DTIA35' = dark meat fish
'DTIA36' = other fish (= *light meat*)
'DTIA37' = shrimp, lobster, scallops

****these variables can be found in the ARIC 'dtia' dataset**

b. Dietary EPA+DHA (total EPA and DHA).
    i. Exclude EPA+DHA from supplements (i.e., do not exclude supplement users, but exclude any portion of EPA+DHA from supplements in these individuals).
    ii. Units: mg/d

'OMEGA' convert from grams to milligrams

'OMEGA' * (0.001) = OMEGAmg (mg/d)

****these variables can be found in the ARIC 'anut2' dataset**

c. For cohorts that have specific information on fried fish: Same exposure as (a), but excluding fried fish.
    i. Units: As in (a), above.

ARIC cannot contribute to this analysis (we did not ask about preparation techniques)

2. **Inclusions**:
    a. ***European descent only (other race groups may be considered at a later date, but such results would be presented in a separate paper)***

3. **Genotype-phenotype association within cohorts:**

    a. Regression models:
        i. Linear regression with robust variance estimators
        ii. Separate models for each outcome

    b. Covariates for adjustment in regression model:
        i. Age in years **v1age01**
        ii. Sex **gender**
        iii. Total energy intake **tcal**
        iv. Study site **center**
        v. Population substructure (e.g, principal components; cohort-specific***, if needed***)

    c. Genetic model: One degree of freedom/ additive model
    The choice of specified allele does not matter, but combining the results will require that the specified allele is stated. See Table 1 below for specifications.

    d. Imputation:
        i. Imputation to Hapmap SNPs
        ii. The 1df model is implemented with imputed estimates of the minor allele count (the 'estimated dosage')

    e. QC methods at cohort levels.
        i. Please see http://depts.washington.edu/chargeco/wiki/QCprocedures

4. **Data Format and Sharing:**

    a. The data delivery format for the meta-analysis will be according to the CHARGE protocol for file sharing.

    b. File names should follow the format :

STUDYNAME_TRAIT_DDMMYY.txt
where TRAIT = FISH or EPADHA

c. ShareSpaces, a secure web-based file-sharing system implemented by the University of Washington's Catalyst computing group, will be used.

d. The following variables should be included when sharing imputed results for meta-analysis **(see below)**. Please note that a README file should be uploaded with a very brief description of the data uploaded, the date, the NCBI human genome reference sequence used (e.g. NCBI 36.2) for strand reference, and the scale of the beta estimates; please also include in the README the SNP HWE p-value, callrate and minor allele frequency filters that have been applied.

variable name	Description
SNPID	SNP ID as rs number
Chr	chromosome number. Use symbols X, XY, Y and mt for non-autosomal markers.
Position	physical position for the reference sequence (build 35 strongly preferred)
coded_all	coded allele, also called modeled allele (in example of A/G SNP in which AA=0, AG=1 and GG=2, the coded allele is G)
noncoded_all	the alternate allele
strand_genome	+ or -, representing either the positive/forward strand or the negative/reverse strand of the human genome reference sequence; to clarify which strand the coded_all and noncoded_all are on
Beta	beta estimate from genotype-phenotype association, at least 5 decimal places -- "NA" if not available
SE	standard error of beta estimate, to at least 5 decimal places -- "NA" if not available
pval	p-value of test statistic, here just as a double check -- "NA" if not available
AF_coded_all	allele frequency for the coded allele -- "NA" if not available
HWE_pval	exact test Hardy-Weinberg equilibrium p-value -- only directly typed SNPs, NA for imputed
Callrate	genotyping callrate after exclusions
n_total	total sample with phenotype and genotype for SNP
imputed	1/0 coding; 1=imputed SNP, 0=if directly typed
used_for_imp	1/0 coding; 1=used for imputation, 0=not used for imputation
oevar_imp	observed divided by expected variance for imputed allele dosage -- NA otherwise

5. **Meta-analysis:**
   a. Fixed Effects
   b. Significance threshold: $p < 1.25 \times 10^{-8}$
   c. Final QC step (e.g. filtering MAF at 1%) at meta-analysis stage
   d. Metabochip cohorts: Include in primary (discovery) meta-analysis.

7.a. Will the data be used for non-CVD analysis in this manuscript? **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**
(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
YES—no overlap

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)?
1656 *"Genome-wide association analysis of macronutrient intake" (Tanaka)*

And 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" (Nettleton)*
1577 *"Interactions between zinc intake and SNPs and their impact on fasting blood glucose levels in multiple cohorts within the CHARGE and MAGIC consortia" (Kanoni)*
1675 *"Low density lipoprotein receptor related protein 1, fatty acids and anthropometric traits" (Smith)*
*#**1716** "CHARGE Nutrition x Gene Working Group: Analysis of interactions between dietary magnesium and SNPs related to fasting glucose and insulin" (McKeown)*

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

GWAS via STAMPEDE & GENEVA, #2006.03

11.b. If yes, is the proposal
_X_ A. primarily the result of an ancillary study (list number* *AS 2006.03*)

ARIC is one of at least 15 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.

Understood, and we will meet this deadline

References

1. Mozaffarian D, Rimm EB. Fish intake, contaminants, and human health: evaluating the risks and the benefits. JAMA. 2006;296:1885-1899.
2. He K, Song Y, Daviglus ML, Liu K, Van Horn L, Dyer AR, Goldbourt U, Greenland P. Fish consumption and incidence of stroke: a meta-analysis of cohort studies. Stroke. 2004;35:1538-1542.
3. Wang C, Harris WS, Chung M, Lichtenstein AH, Balk EM, Kupelnick B, Jordan HS, Lau J. n-3 Fatty acids from fish or fish-oil supplements, but not {alpha}-linolenic acid, benefit cardiovascular disease outcomes in primary- and secondary-prevention studies: a systematic review. Am J Clin Nutr. 2006;84:5-17.
4. Eslick GD, Howe PR, Smith C, Priest R, Bensoussan A. Benefits of fish oil supplementation in hyperlipidemia: a systematic review and meta-analysis. Int J Cardiol. 2009;136:4-16.
5. Geleijnse JM, Giltay EJ, Grobbee DE, Donders AR, Kok FJ. Blood pressure response to fish oil supplementation: metaregression analysis of randomized trials. J Hypertens. 2002;20:1493-1499.
6. Mozaffarian D, Geelen A, Brouwer IA, Geleijnse JM, Zock PL, Katan MB. Effect of fish oil on heart rate in humans: A meta-analysis of randomized controlled trials. Circulation. 2005;112:1945-1952.
7. Mozaffarian D. UpToDate: Fish oil and marine omega-3 fatty acids. Available at: http://www.uptodate.com/patients/content/topic.do?topicKey=~P22PNBTmTumdat. Accessed October 13, 2010.
8. Reed DR, Bachmanov AA, Beauchamp GK, Tordoff MG, Price RA. Heritable variation in food preferences and their contribution to obesity. Behav Genet. 1997;27:373-387.
9. Hasselbalch AL, Heitmann BL, Kyvik KO, Sorensen TI. Studies of twins indicate that genetics influence dietary intake. J Nutr. 2008;138:2406-2412.