ARIC Manuscript Proposal # 1738

PC Reviewed: 1/11/11   Status: A   Priority: 2
SC Reviewed: _________   Status: _____   Priority: ____

1.a. Full Title: Interaction between a multi-factorial diet score and genetic loci for fasting glucose and insulin

b. Abbreviated Title (Length 26 characters): gene loci x diet score

2. Writing Group:
Writing group members: Tentative list below; additional authors will be added; other ARIC investigators are welcome

<table>
<thead>
<tr>
<th>PARTICIPATING COHORT</th>
<th>N</th>
<th>PHENOTYPIC ANALYST</th>
<th>MAIN ANALYST</th>
<th>AUTHOR (cohort lead)</th>
<th>AUTHOR 2</th>
<th>AUTHOR 3</th>
<th>AUTHOR 4</th>
<th>AUTHOR 5</th>
<th>AUTHOR (cohort senior)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIC</td>
<td>7,201</td>
<td>na</td>
<td>Jennifer Nettleton</td>
<td>Jennifer Nettleton</td>
<td>Keri Monda</td>
<td>Kari North</td>
<td>Jim Pankow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MESA</td>
<td>2,271</td>
<td>Jennifer Nettleton</td>
<td>Ani Manichaikul</td>
<td>David Jacobs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Framingham</td>
<td>5,838</td>
<td>Nicola McKeeown</td>
<td>Julius Ngwa</td>
<td>Nicola McKeeown</td>
<td>Julius Ngwa</td>
<td>Josee Dupuis</td>
<td>Adrienne Cuppies</td>
<td>James Meigs</td>
<td></td>
</tr>
<tr>
<td>CHS</td>
<td>2,800</td>
<td>Dariush Mozaffarian</td>
<td>Rozenn Lemaitre</td>
<td>Rozenn Lemaitre</td>
<td>Dariush Mozaffarian</td>
<td>Kenneth Mukamal</td>
<td>Lue Djuusse</td>
<td>Kerri Wiggins</td>
<td>David Siscovick</td>
</tr>
<tr>
<td>Family HS</td>
<td>3,523</td>
<td>na</td>
<td>Mary Wojczynski</td>
<td>Mary Wojczynski</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health ABC</td>
<td>1,497</td>
<td>Denise Houston</td>
<td>Denise Houston</td>
<td>Kurt Lohman</td>
<td>Yongmei Liu</td>
<td></td>
<td></td>
<td>Steve Kritchevsky</td>
<td></td>
</tr>
<tr>
<td>InCHIANTI</td>
<td>1,071</td>
<td>na</td>
<td>Toshiko Tanaka</td>
<td>Toshiko Tanaka</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHRAS</td>
<td>960</td>
<td>na</td>
<td>Stavroula Kanoni</td>
<td>Stavroula Kanoni</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>George Dedoussis</td>
</tr>
<tr>
<td>GENDAI</td>
<td>1,123</td>
<td>na</td>
<td>Ioanna Ntalia</td>
<td>Ioanna Ntalia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YOUNG FINNS</td>
<td>800</td>
<td>Vera Mikkila</td>
<td>Vera Mikkila</td>
<td>Terho Lehtimaki</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>Olli Raitakari</td>
<td></td>
</tr>
<tr>
<td>ULSAM</td>
<td>933</td>
<td>Per Sjogren</td>
<td>Erika Ingelsson</td>
<td>Per Sjogren</td>
<td>Erika Ax</td>
<td>TBD</td>
<td>--</td>
<td>--</td>
<td>Erik Ingelsson</td>
</tr>
<tr>
<td>GLACIER</td>
<td>21,638</td>
<td>Ingegerd Johansson</td>
<td>Paul Franks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Paul Franks</td>
</tr>
<tr>
<td>ROTTERD AM</td>
<td>2,305</td>
<td>Jessica Kieffe</td>
<td>Frank van Rooij</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALMO</td>
<td>3,679</td>
<td>na</td>
<td>Emily Sonestedt</td>
<td>Emily Sonestedt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Marja Ortho-Melander</td>
</tr>
</tbody>
</table>

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

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:  

3. Timeline: Cohort-specific data analyses: February 1, 2011  
Meta-analysis: March 15, 2011  
Manuscript drafting complete: May 15, 2011

4. Rationale:  
Dietary pattern research has been an important focus in nutritional epidemiology in recent decades. Investigators have used both empirical (data-driven) and a priori (investigator-defined) methods to globally characterize food intake, acknowledging the fact that neither single nutrients nor individual foods are consumed in isolation. There are several unifying characteristics of these global scores (both those derived empirically and those investigator-defined) that have been associated with chronic diseases and their risk factors. While there are disease-specific (cardiovascular disease, diabetes, cancer) and sample-specific (region, age, race/ethnicity) nuances that distinguish dietary patterns, these differences tend to be subtle. Diets that emphasize plant foods (e.g., whole grains, fruits, vegetables, nuts and seeds) and plant and marine sources of fat (e.g., nuts and seeds and fatty fish) and de-emphasize red meat, foods high in sugar and salt, and highly refined grains are associated with lower prevalence and incidence of chronic diseases and their risk factors.

Associations between these dietary pattern scores and health outcomes are often attributed to their nutrient constituents. For example, “heart healthy” or “prudent” dietary patterns are lauded for their fiber, antioxidant micronutrient, and unsaturated fat contributions. However, one of the arguments in support of the more global dietary pattern approach in nutrition research is that nutrient components are likely to be interactive or synergistic and nutrients often show high intercorrelations due to shared food sources. Thus, isolating effects to a single food constituent is likely to be both a confounded and possibly inaccurate representation of the true nature of diet. While in some respects it does make sense to approach questions about biological mechanisms by focusing at the level of nutrient, the biological concentration of a given nutrient at the tissue of interest is influenced by numerous factors, including the food matrix in which the nutrient was delivered as well as the context of other nutrients present in the diet. Thus, characterization of the overall dietary pattern may better
characterize some of these factors and may be particularly valuable in the absence of biomarkers or tissue-specific measures of nutrient concentrations.

Data from numerous well-controlled dietary intervention studies suggest that individuals respond variably to dietary interventions, and some of this variability is thought to be related to genetic factors, i.e., interaction between dietary intake and genotype. Historically, investigations focused on characterizing interactions between diet and genetic factors have focused on nutrients as the biologically active representatives of “diet,” based on the premise that genetic variations manifest in altered biological pathways whose resultant effects on human health could be magnified or minimized by the intake of specific nutrients. However, in the context of epidemiological data (vs. controlled, randomized interventions), perhaps the more prudent question is one focused on the overall pattern of food consumption and its interaction with genetic factors. This level of the dietary hierarchy is also particular appealing to public health research, where conclusions about which foods to eat are more “digestible” to the general public than conclusions about which and how much of ‘X’ nutrient to consume.

Science is moving closer to being able to estimate an individual’s risk for common chronic diseases based on genotype at specific loci, thus, increasing the imperative to identify modifiable lifestyle factors, like diet, that can effectively delay or prevent the onset of these seemingly inevitable “diagnoses.” Whether an evidence-based dietary regimen that is consistent with population-level guidelines and prior research can prevent early onset of disease and/or contribute to risk factor management is uncertain. While results from the Diabetes Prevention Program do support this supposition, additional data from other larger population-based studies with a singular focus on diet is needed. References on page 8.

5. Main Hypothesis/Study Questions:

Project Aims:

(Preparatory Aim)

- Create a Diet Score reflecting “dietary guidelines aimed to main health” that can be applied across multiple, large cohort studies spanning regions of the U.S., Northern Europe, and Mediterranean Europe.

To create the Diet Score we evaluated the following:

- Country-specific dietary guidelines
- Results of investigations of the associations between specific dietary factors and patterns with respect to diabetes and its risk factors
• Data quality and availability in participating cohorts

• Regional food usage patterns

After reviewing the above, we decided to create a Diet Score that reflects the intake of 5 “healthy” food groups [whole grains, fruit, vegetables, fish, nuts/seeds] and 4 “unhealthy” food/beverage groups [red and processed meat, sugar-sweetened beverages, desserts/sweets, fried potatoes]

(details described below in methods/analytic plan).

(Research Aims)

• Determine the magnitude of the association of the Diet Score with fasting glucose and fasting insulin. (Diet Score main “effect”)

• Determine whether a higher Diet Score mitigates the glucose- or insulin-raising effect of the SNPs identified by the MAGIC consortium as significant predictors of fasting glucose or insulin in individuals without diabetes. (Diet Score x SNP interaction)
  o Determine whether a higher Diet Score mitigates glucose-raising effect of overall genetic burden at the 16 loci identified by the MAGIC consortium as predictive of high fasting glucose (Diet Score x Genetic Risk Score interaction—details described below in methods/analytic plan)

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 & Analysis Plan

Diet Score:

‘HEALTHY’ FOOD GROUPS (quartile ranks summed to generate diet score: Qt 1 = 0 pts; Qt 2 = 1 pt; Qt 3 = 2 pts; Qt 4 = 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 (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: Qt 4 = 0 pts; Qt 3 = 1 pt; Qt 2 = 2 pts; Qt 1 = 3 pts & then summed to generate diet score)
  o Red Meat & Processed Meat (combined group*)
  o Sugar-sweetened beverages (soda pop and sugar-sweetened, artificially fruit flavored juices)
  o Fried Potatoes* (in cohorts where intake was quantified)
  o Desserts & Sweets

Score maximum = 27; Score minimum = 0
Genetic Risk: **
- 16 SNPs identified as significant predictors of high fasting glucose in the MAGIC consortium \(^\text{10}\)
- 2 SNPs identified as significant predictors of high fasting insulin in the MAGIC consortium \(^\text{10}\)
- (glucose) Genetic Risk Score\(^\dagger\): number of glucose-raising alleles summed across 16 glucose-related loci

\[^{\dagger}\text{Note: one SNP was predictive of both fasting glucose and insulin (thus, total number of SNPs = 17)}\]

Initially, we plan to use an un-weighted glucose-GRS; we will consider options for weighting the score after reviewing our data

<table>
<thead>
<tr>
<th>GlucoseSNP</th>
<th>Chr</th>
<th>Nearest gene</th>
<th>Effect Allele</th>
<th>Other Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs340874</td>
<td>1</td>
<td>PROX1</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>rs560887</td>
<td>2</td>
<td>G6PC2</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>rs780094</td>
<td>2</td>
<td>GCKR</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>rs11708067</td>
<td>3</td>
<td>ADCY5</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>rs11920090</td>
<td>3</td>
<td>SLC2A2</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>rs4607517</td>
<td>7</td>
<td>GCK</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>rs2191349</td>
<td>7</td>
<td>DGKB/TMEM195</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>rs11558471</td>
<td>8</td>
<td>SLC30A8</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>rs7034200</td>
<td>9</td>
<td>GLIS3</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>rs10885122</td>
<td>10</td>
<td>ADRA2A</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>rs4506565</td>
<td>10</td>
<td>TCF7L2</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>rs10830963</td>
<td>11</td>
<td>MTNR1B</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>rs7944584</td>
<td>11</td>
<td>MADD</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>rs11605924</td>
<td>11</td>
<td>CRY2</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>rs174550</td>
<td>11</td>
<td>FADS1</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>rs11071657</td>
<td>15</td>
<td>FAM148B</td>
<td>A</td>
<td>G</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>InsulinSNP</th>
<th>Chr</th>
<th>Nearest gene</th>
<th>Effect Allele</th>
<th>Other Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs780094*</td>
<td>2</td>
<td>GCKR</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>rs35767</td>
<td>12</td>
<td>IGF1</td>
<td>G</td>
<td>A</td>
</tr>
</tbody>
</table>

*note that this SNP was also associated with fasting glucose in MAGIC*

Exclusions:
- Prevalent type 2 diabetes
  - Self-reported diabetes
  - Taking medication for diabetes
  - Fasting glucose ≥ 126 mg/dL (≥ 7 mmol/L)
- Non-fasting status
- Implausible dietary data
  - Cohort-specific definition
- Non-white race

Dependent Variables:
1. Fasting glucose concentration (mmol/L, continuous, untransformed)
2. Fasting insulin concentration (pmol/L, continuous, with natural log transformation)
Diet Score-Fasting Glucose/Insulin ("main effects") Analysis:

Model 1:
- sex
- age (continuous)
- field center (if needed)
- total energy intake (kcal intake per day, continuous)

Model 2:
- model 1 covariates
- education level (cohort-specific definition)
- smoking (cohort-specific definition)
- physical activity (cohort-specific definition)
- alcohol intake (grams of ethanol per day, continuous)

Model 3:
- model 2 covariates
- Body Mass Index (kg/m², continuous)

Diet Score x SNPs** Interaction Analysis:

Model Covariates:
- sex
- age (continuous)
- field center (if needed)
- total energy intake (kcal intake per day, continuous)
- population substructure adjustment as needed

1. Multivariable INTERACTION Model → GLUCOSE:
   fasting glucose = SNP (estimated copies of risk allele), DIET SCORE (continuous), SNP*DIET SCORE + model covariates listed above

2. Multivariable INTERACTION Model → INSULIN:
   fasting (ln)insulin = SNP (estimated copies of risk allele), DIET SCORE (continuous), SNP*DIET SCORE + model covariates listed above

Data Sharing:

I → From the 3 hierarchical models specified above, each cohort will provide beta regression coefficient, SE, and p value for the multivariable-adjusted DIET SCORE term

II → Using the single interaction model specified above, each cohort will provide beta regression coefficient, SE, and p value for
   A) DIET SCORE*SNP product term
   B) DIET SCORE marginal term
   C) SNP marginal term
   D) intercept

Also, for descriptive purposes, please provide the following:
- Cohort-specific sex distribution (%female)
- Cohort-specific age (years mean ± SE)
- Cohort-specific energy intake (kcal/day mean ± SE)
- Cohort-specific fasting glucose concentrations (mmol/L mean ± SE)
- Cohort-specific fasting insulin concentrations (pmol/L mean ± SE)—values before and after ln-transformation
Meta analyses:
Meta analyses will be conducted on the regression coefficients for

I. The DIET SCORE association with fasting glucose (3 models) and fasting insulin (3 models)

II. The DIET SCORE x SNP interaction term (1 model) for interaction with each of 16 glucose SNPs and 2 insulin SNPs for both fasting glucose and fasting insulin outcomes

III. The DIET SCORE x GENETIC RISK SCORE interaction term (1 model) for the fasting glucose outcome

Significance level of tests of associations between Diet Score & Fasting Glucose and Insulin: \( p < 0.05 \)

Significance level of tests of interaction: (Bonferroni correction)
\( p \leq 0.0025 = 0.05/20 \): 18 diet score x SNP tests (16 for glucose; 2 for insulin) + 2 diet score x genetic risk score test (1 for glucose; 1 for insulin)

7.a. Will the data be used for non-CVD analysis in this manuscript?

*Fasting glucose is the primary outcome*

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.csec.unc.edu/ARIC/search.php](http://www.csec.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)?

Other 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”
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 and 2006.03)
ARIC is one of 14 cohort studies contributing data to the CHARGE/MAGIC-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.

REFERENCES