ARIC Manuscript Proposal # 1370

<table>
<thead>
<tr>
<th>PC Reviewed: 05/13/08</th>
<th>Status: A</th>
<th>Priority: 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC Reviewed: _________</td>
<td>Status: ____</td>
<td>Priority: ____</td>
</tr>
</tbody>
</table>

1a. **Full Title:** Analysis of gene-environment interactions: SNPs from adiposity GWAS and physical activity

b. **Abbreviated Title:** G-E GWAS SNPs and activity

2. **Writing Group:**
Writing group members:
- Keri Monda
- Kari North
- Eric Boerwinkle
- Ellen Demerath
- Linda Kao
- Braxton Mitchell (with the OOA Study)
- Caroline Fox (with the FHS)
- Tamara Harris (with the AGES Study)

**Other investigators welcome**

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

First author: Keri Monda
Address: Department of Epidemiology
University of North Carolina at Chapel Hill
137 E. Franklin St, Suite 306
CB #8050
Chapel Hill, NC 27514
Phone: 919-966-1403 Fax: 919-966-9800
E-mail: monda@unc.edu

Corresponding/senior author (if different from first author correspondence will be sent to both the first author & the corresponding author): Kari North
Address: Department of Epidemiology
University of North Carolina at Chapel Hill
137 E. Franklin St, Suite 306
CB #8050
Chapel Hill, NC 27514
Phone: 919-966-2148 Fax: 919-966-9800
E-mail: kari_north@unc.edu
3. Timeline:
   - Statistical analyses: August – October, 2008
   - Manuscript revision: January 2009
   - Manuscript submission: February 2009

4. Rationale:
Several lines of evidence support the role of genetics in the regulation of body mass, including longitudinal family and twin studies which show that BMI, weight, and weight change are all heritable traits (Adams, Hunt et al. 1993; Austin, Friedlander et al. 1997; Lee, Reed et al. 1997; Bouchard, Perusse et al. 1998; Comuzzie and Allison 1998; Hunt, Katzmarzyk et al. 2002; Loos and Bouchard 2003). However, most forms of obesity do not follow simple Mendelian modes of inheritance and thus investigating potential genetic variants that contribute to common forms of obesity will require large population-based studies. Linkage analyses of family-based data have identified areas of the human genome that are associated with adiposity traits (Golla, Strauch et al. 2003; Fox, Heard-Costa et al. 2005). In fact, according to the annually updated ‘Obesity Gene Map’ (Rankinen, Zuberi et al. 2006) 253 quantitative trait locus (QTL) regions for obesity-related phenotypes have been identified in 61 genome-wide scans, and a total of 52 genomic regions that harbor QTLs replicated in two or more studies. Despite this, no specific genetic variants clearly responsible for any of the linkage signals have been identified. It is only with recent major technological advances that we have rapidly expanded options for the evaluation of genetic variation at the level of the single nucleotide polymorphism (SNP).

Genome-wide Association (GWA) studies interrogate whether variation across the human genome in the form of SNPs is associated with given phenotypes. GWAS are now widely recognized as powerful data-driven tools for identifying genetic variants related to common complex diseases such as obesity. GWA was used to identify variants near \textit{INSIG2} reported to be associated with obesity in the Framingham Heart Study and replicated in four of five studies of adults and children (Herbert, Gerry et al. 2006). More recently, a widely-replicated result between \textit{FTO} and obesity has been reported (Dina, Meyre et al. 2007; Frayling, Timpson et al. 2007; Scuteri, Sanna et al. 2007). Replication of findings is a key ingredient in genetic epidemiology studies and investigators are encouraged to set up collaborations to facilitate this. Failure to replicate could be due to many reasons including sample differences, lack of power to find an effect, incomplete phenotype harmonization, among others. Moreover, genetic effects on obesity phenotypes are likely to be strongly influenced by the environment.

Studies suggest that genotype may influence sensitivity of individuals to environmental stressors (Plomin, DeFries et al. 1977; Bray 2000; Chakravarti and Little 2003). The well-known ‘thrifty gene’ hypothesis (Neel 1962; Neel 1999) argues that genes favoring minimum energy expenditure and maximum energy storage were preferentially selected because of their ability to provide an advantage to populations that frequently experienced starvation by allowing for excess adipose storage when food was plentiful, and provides one explanation of the human response to the modern environment where...
the food supply is constant throughout the year and the energy demands of daily work have greatly decreased. On an individual level, obesity remains a very heterogeneous disease, and individuals’ phenotypic responses differ greatly when exposed to the same environmental influences. The term gene-environment interaction refers to the idea that one’s genotype may influence how he or she responds to the effects of the environment (Perusse and Bouchard 1999), and in its absence, the phenotypic response to an environmental effect is similar across genotypes.

There is an extensive literature devoted to the study of the effects of physical activity (PA) on obesity, and it is well established that there is great interindividual variation in response to exercise. This interindividual variability in response to lifestyle change is likely to be partly determined by genetics and provides a rationale for studying genes and environmental factors simultaneously. In a large French cohort, significant associations were noted between body weight, BMI and waist and hip circumferences and the ADRB2 Gln27Glu polymorphism, but the associations were limited to sedentary subjects, not in the physically active (Meirhaeghe, Helbecque et al. 1999). Similarly, in ARIC, a significant interaction between the GNB3 825C>T polymorphism and physical activity was found in predicting obesity status in African Americans where the T allele was associated with lower prevalence of obesity in active individuals, and a higher prevalence of obesity in sedentary individuals (Grove, Morrison et al. 2007). Nonetheless, even though most researchers agree that the development of obesity is dependent upon the presence of not only specific genetic factors but also certain environmental conditions, investigations are uncommon and there is a great need for large samples with documented environmental exposure data, like those available in ARIC, to investigate potential gene-environment interaction.

We have submitted an earlier manuscript proposal (first author Kari North – MS #1368) for the initial analyses on the genome-wide SNP data (~1,000,000 SNPs) available on the ARIC sample through its collaboration with the Broad Institute. Phenotypes assessed included the following:

- Baseline BMI and waist circumference measures
- Mean and maximum BMI and waist circumference measurements
- Height
- Alternative measures of adiposity (skinfold measures)
- Weight and/or BMI change

We plan also to investigate qualitative traits based on the quantitative traits listed above. For example, using standard cutoffs, we will investigate the phenotypes overweight (BMI≥25 kg/m²), obesity (BMI≥30 kg/m²), and high waist circumference (≥88 cm women, ≥102 cm men).

5. **Main Hypotheses/Study Questions:**
   1) To test the interaction effect between SNPs identified through the genome wide association study (GWAS) as significantly associated with adiposity traits and baseline physical activity level on baseline adiposity traits (see list above).
2) To test the interaction effect between identified SNPs and physical activity level on longitudinal changes in adiposity traits

6. Design and Analysis:
Subjects and Sample size:
The usual DNA consent restriction and missing data exclusion criteria will be used.
Mean (SD) levels and sample sizes by race for the main phenotypes to be assessed are in the table below.

<table>
<thead>
<tr>
<th></th>
<th>Whites</th>
<th>African Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Body mass index (kg/m²)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>27.0 (4.9) n=11468</td>
<td>29.6 (6.2) n=4196</td>
</tr>
<tr>
<td>v2</td>
<td>27.3 (4.9 ) n=10720</td>
<td>30.0 (6.3) n=3509</td>
</tr>
<tr>
<td>v3</td>
<td>27.9 (5.2) n=9838</td>
<td>30.4 (6.4) n=2951</td>
</tr>
<tr>
<td>v4</td>
<td>28.3 (5.2) n=8946</td>
<td>30.6 (6.4) n=2603</td>
</tr>
<tr>
<td><strong>Waist circumference (cm)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>96.3 (13.4) n=11464</td>
<td>99.2 (15.2) n=4198</td>
</tr>
<tr>
<td>v2</td>
<td>97.0 (13.9) n=10722</td>
<td>101.0 (15.2) n=3517</td>
</tr>
<tr>
<td>v3</td>
<td>100.0 (13.9) n=9841</td>
<td>102.8 (15.7) n=2948</td>
</tr>
<tr>
<td>v4</td>
<td>101.4 (14.1) n=8951</td>
<td>104.1 (15.8) n=2603</td>
</tr>
<tr>
<td><strong>Skinfold (mm)</strong>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>46.4 (16.7) n=11436</td>
<td>59.5 (25.1) n=4179</td>
</tr>
<tr>
<td>v2</td>
<td>42.2 (15.4) n=10723</td>
<td>51.8 (22.2) n=3513</td>
</tr>
<tr>
<td><strong>Sport PA (Baecke index units)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>2.5 (0.8) n=11438</td>
<td>2.1 (0.7) n=4185</td>
</tr>
<tr>
<td>v3</td>
<td>2.6 (0.8) n=9805</td>
<td>2.3 (0.7) n=2900</td>
</tr>
<tr>
<td><strong>Leisure PA (Baecke index units)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>2.5 (0.5) n=11462</td>
<td>2.1 (0.6) n=4190</td>
</tr>
<tr>
<td>v3</td>
<td>2.4 (0.5) n=9839</td>
<td>2.1 (0.6) n=2908</td>
</tr>
<tr>
<td><strong>Work PA (Baecke index units)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>2.6 (0.7) n=8310</td>
<td>2.8 (0.7) n=3038</td>
</tr>
<tr>
<td>v3</td>
<td>2.4 (0.8) n=5913</td>
<td>2.5 (0.8) n=1867</td>
</tr>
<tr>
<td><strong>Total sport PA (METmins/wk)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>84.0 (64.9) n=7980</td>
<td>67.6 (60.0) n=1817</td>
</tr>
<tr>
<td>v3</td>
<td>86.2 (63.5) n=6735</td>
<td>73.1 (59.6) n=1499</td>
</tr>
<tr>
<td><strong>Mod/vig sport PA (METmins/wk)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>82.1 (64.8) n=7980</td>
<td>64.4 (59.5) n=1818</td>
</tr>
<tr>
<td>v3</td>
<td>85.1 (63.3) n=6735</td>
<td>70.7 (59.2) n=1500</td>
</tr>
<tr>
<td><strong>Vig sport PA (METmins/wk)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>19.9 (47.1) n=7981</td>
<td>16.2 (46.3) n=1818</td>
</tr>
<tr>
<td>v3</td>
<td>16.2 (43.4) n=6736</td>
<td>11.1 (40.3) n=1506</td>
</tr>
</tbody>
</table>

*Sum of triceps and subscapular skinfolds

Publication strategy
With the existing collaboration we have with the Framingham Heart Study (FHS), studies of the Old Order Amish (OOA), and the Age Gene/Environment Susceptibility (AGES)-Reykjavik Study (investigators’ names included in the writing group), we will seek to
replicate any significant results we might find in this work. We recognize that physical activity has been measured differently in these populations and we will seek to harmonize environmental characterization as best as possible. Brief descriptions of these studies are below. This group will facilitate replication for findings in the white sample; we are still working to set up collaborations to replicate any findings from the African American sample.

FHS: The Framingham Heart Study is a longitudinal study of cardiovascular disease with the original participants (aged 30-62) recruited in 1948. A second-generation group composed of the original participants’ adult children and their spouses was enrolled in 1971. Finally, a third-generation was enrolled in 2002. Genotyping data is available via the Affymetrix 500K SNP chip on a sample size of approximately 9300 individuals. The FHS is a joint project of the NHLBI and Boston University.

OOA: Approximately 1,600 Old Order Amish individuals from Lancaster, PA will also be included. These individuals, aged 18 years and older, were participants of studies carried out by investigators at the University of Maryland, Baltimore (the HAPI Heart Study, the Amish Family Calcification Study, and the Amish Longevity Study). Genotyping data (via the Affymetrix 500K or 1M SNP chip) is available on approximately 1600 individuals.

AGES-Reykjavik: The Age Gene/Environment Susceptibility-Reykjavik Study was initiated in 2002 to examine risk factors in relation to disease and disability in old age. The sample is drawn from an established population-based cohort, the Reykjavik Study, of men and women born between 1907 and 1935 in Iceland. Genotyping data will be available in June via the Affymetrix chip.

Definitions and treatment of variables
Genotype: For these analyses we will likely employ a codominant model when sample size permits (cell size > 10). This model allows for flexibility of effect and has been shown to have good power to find an effect when the underlying mode of inheritance is unknown (Lettre, Lange et al. 2007). In a codominant model two indicator variables are entered in the model with the major homozygote used as the referent category. This results in a regression coefficient for the heterozygote and the minor homozygote; an overall (or global) p-value is used to assess the statistical significance of the result.

Physical activity (PA): In ARIC, physical activity was measured at visits one and three utilizing a modified Baecke Questionnaire of Habitual Physical Activity resulting in three indices of activity: sport activity, work activity, and leisure activity. We propose to primarily examine gene-environment interaction using the sport index because it has been shown to provide the most reliable and valid results. Nonetheless, we will also examine interactions using the work and leisure indices as well as a summary measure of total activity. Physical activity indices will be examined as continuous measures (thus increasing power to detect interaction) as well as categorized based on tertiles or other more appropriate (data-driven) categorizations. We will also examine variables utilizing metabolic equivalents (METmins/week for total PA, moderate & vigorous PA, and
vigorous PA) that were derived from the Baecke questionnaire and which further interrogate intensity.

Phenotype measures: Measures used in phenotype characterization in ARIC (BMI, waist circumference, height, and sum of triceps and subscapular subcutaneous skinfolds) will be primarily defined as quantitative traits. Normality will be assessed prior to analysis. As stated above, the qualitative traits overweight (BMI ≥ 25 kg/m²), obesity (BMI ≥ 30 kg/m²), and high waist circumference (≥ 88 cm men, ≥ 102 cm women) will be created from their quantitative versions.

Analysis strategy / statistical analysis
SNPs found through the GWAS will be tested for interaction with physical activity if they exhibit at least moderate main effects in the individual models. We will assess SNPs found through GWAS if they have an uncorrected p-value < 0.05. SNPs found through other GWA studies that do not appear in the ARIC data will also be evaluated assuming that there is no overlap in existing manuscript proposals (e.g. Demerath’s proposal with FTO). We chose the fairly liberal uncorrected p-value of <0.05 in order to minimize false negatives given that obesity is such a heterogeneous disease and given that we know that the environment (i.e. PA) has a major effect on the distribution of the phenotype in the population. Because it is likely that we will be doing a large number of statistical tests which increase the likelihood of a false positive result, we will need to correct for multiple testing. We will do this both by using the Bonferroni correction, which is often criticized for being overly-conservative, as well as the false discovery rate (FDR) method (Benjamini, Drai et al. 2001; van den Oord and Sullivan 2003).

General linear models for quantitative adiposity traits and logistic models for qualitative traits will include the main effects of SNP and PA as well as an interaction term. Significant association of interaction will be assessed by means of a likelihood ratio test (p < 0.10 due to low power of test). We will test both minimally adjusted models (age, sex, field center) and more extensively adjusted models. Possible additional covariates include smoking, alcohol intake, and education level.

Analyses stratified by PA level will also be pursued for candidates with interaction terms near statistical significance (e.g. 2.0 > p > 1.0) and may be important in addition to formal testing of interactions. For example, if individuals with low levels of physical activity are more susceptible, there may be more power to find genome wide association in this subgroup than to formally test for an interaction with physical activity.

As a secondary aim, we will examine the interaction of baseline PA with genotype on changes in adiposity measures. Change in adiposity will be assessed using difference models for individuals with >1 observation (differences will be divided by the number of years between the observations to create an annualized change measure). We may also employ mixed effects models to incorporate all repeated observations, thus taking advantage of the vast majority of the data. These models will include sex, age, field center, genotype, physical activity, time, and a genotype*PA interaction term.
Alternative ways to model physical activity include looking at average PA across the two exams (visits 1 and 3) in order to minimize errors associated with measuring PA. Additionally, we can evaluate the interaction between change in PA and genotype on change in phenotype to assess whether the relationship between change in PA on change in phenotype differs by genotype, with the caveat that change models are notoriously difficult due to compounding error.

Because systematic differences in ancestry can produce spurious associations, all analyses will be stratified by race to account for systematic allele frequency differences between racial groups. However, we will also need to account for population substructure within racial groups. While there are a number of methods available to the analyst, we can take advantage of GWAS data most effectively using the principal components analysis method developed by Price and colleagues (Price, Patterson et al. 2006), implemented in the software EIGENSOFT. This method explicitly models ancestry and has higher power to detect true associations than other methods. Principal component factor scores will be incorporated into genetic models to account for population stratification in each of the samples.

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 ICTDER02 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 ICTDER02 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 ICTDER02 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
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)?

#1358 (Demerath): “Interaction between FTO genotype and physical activity level on adiposity: The Atherosclerosis Risk in Communities (ARIC) Study”

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 __AS#1995.07__) 
___ 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/](http://www.cscc.unc.edu/aric/forms/)

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.

References


