1a. Full Title: Genome-wide interaction analysis of SNPs and physical activity on obesity traits

b. Abbreviated Title: Obesity gene-by-PA GWAS interaction

2. Writing Group:
   Keri Monda
   Kari North
   Eric Boerwinkle
   Ellen Demerath
   Collaborators with the CHARGE consortium
   Collaborators with the GIANT consortium
   Collaborators with the CARE consortium

   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: January - February, 2011
   Manuscript preparation: March - June, 2011
   Manuscript revision: July - August 2011
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 body mass index (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 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 most recently 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. Obesity researchers have recently had notable success using GWAS in discovering genetic variants for the anthropometric traits BMI, waist circumference (WC) and waist-hip ratio (WHR) (Frayling, Timpson et al. 2007; Loos, Lindgren et al. 2008; Heard-Costa, Zillikens et al. 2009; Thorleifsson, Walters et al. 2009; Willer, Speliotes et al. 2009; Heid, Jackson et al.; Speliotes, Willer et al.). The vast majority of these variants were identified by the efforts of multiple cohorts collaborating to form consortia.

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 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). More recently, a
number of studies have been published investigating the interaction of the FTO gene and
adiposity phenotypes (Andreasen, Stender-Petersen et al. 2008; Rampersaud, Mitchell et al.
2008; Vimaleswaran, Li et al. 2009). However, no studies have been published that investigate
gene-by-physical activity interaction on a genome-wide scale, and studies that are limited to the
investigation of candidate genes discovered through main-effects GWAS are limited in that
interaction effects may be masked in analyses not previously designed to look at stratum-specific
effects. Thus there is a great need for large samples with documented environmental exposure
data, like those available in ARIC, to investigate gene-environment interaction on a genome-
wide level.

Note: For analyses of individuals of European-descent, these analyses will be done in
collaboration with the CHARGE and GIANT consortia. For analyses of individuals of African-
American descent, analyses will be done in collaboration with the CARE consortium
(manuscript proposal currently under review).

5. Main Hypotheses/Study Questions:
To test the interaction effect genome-wide between SNPs and baseline physical activity level
on baseline adiposity traits (BMI, WC, and WHR).

6. Design and Analysis:
Subjects and Sample size:
Individuals with available anthropometric, exposure, and covariate measures. The usual DNA
consent restriction and missing data exclusion criteria will be used. Use of GWAS data in
African-Americans will follow CARE procedures (see above) and will likely be published as a
separate paper.

Definitions and treatment of variables
Genotype: Genome-wide genotyping data (~900K SNPs) and available imputed data (~2.8 M
SNPs) from the freeze 3 ARIC data. Standard exclusion criteria will be applied.

Physical activity (PA): In ARIC, PA 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 as well as dichotomized based on more appropriate (data-driven)
categorizations. We may also examine variables utilizing metabolic equivalents (METs/week for total PA, moderate & vigorous PA, and vigorous PA) derived from the Baecke questionnaire and which further interrogate intensity.

Phenotype measures: BMI, waist circumference (WC), and waist-hip ratio (WHR) will be defined as quantitative traits. Outcome variables will be transformed as necessary.

Covariates: Models will be minimally adjusted for age, sex, field center, and smoking status. We will also test potential confounders such as educational status and alcohol intake during model building. Principal components will be controlled for in models to account for population substructure.

Analysis strategy / statistical analysis
Additive models will be used to estimate the interaction between SNPs and PA on adiposity traits. We will run models that (1) contain both the main effects of SNP and continuous PA and the interaction term for SNP*PA as well as (2) models stratified by dichotomized PA level. We will seek replication of results within our existing collaboration with the CHARGE and CARE Consortia. Only summary data including beta coefficients and p-values as well as other necessary data (strand, etc) will be shared with collaborators. No individual-level data will be shared. Meta-analyses based on both effect estimates as well as p-values will be run.

Phenotype harmonization: We recognize the harmonization of PA data between CHARGE/GIANT and CARE cohorts is an important and complex issue. We are currently working with collaborators on the best way to accomplish this.

Multiple testing: We will control for multiple testing using the Bonferroni correction on an overall alpha=0.05.

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
   ___x_ 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  
__x__ Yes  
_____ 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)?  
#1358 (Demerath): “Interaction between FTO genotype and physical activity level on adiposity: The Atherosclerosis Risk in Communities (ARIC) Study” 
#1370 (Monda): “Analysis of gene-environment interactions: SNPs from adiposity GWAS and physical activity.”

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  
_x__ A. primarily the result of an ancillary study (AS #2006.03 & 2007.02)  
___ 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/

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.

Literature cited:


